



Achieve, Inc.

Measuring Up 2005

A Report on Assessment Anchors
and Tests in Reading and
Mathematics for

PENNSYLVANIA

ACHIEVE'S
BENCHMARKING
INITIATIVE



Achieve, Inc.

Created by the nation's governors and business leaders, Achieve, Inc., is a bipartisan, non-profit organization that helps states raise academic standards, improve assessments and strengthen accountability to prepare all young people for postsecondary education, work and citizenship. Achieve has helped nearly half the states benchmark their standards and tests against the best examples in this country and abroad and work in partnership to improve teaching and learning. Achieve serves as a significant national voice for quality in standards-based reform and regularly convenes governors, CEOs and other influential leaders at National Education Summits and other gatherings to sustain support for higher standards and achievement for all of America's schoolchildren.

Achieve helps states raise academic standards, measure performance against those standards, establish clear accountability for results and strengthen public confidence in our education system. To do this, we:

- help states **benchmark** their standards, assessments and accountability systems against the best in the country and the world;
- provide sustained **public leadership** and advocacy for the movement to raise standards and improve student performance;
- build **partnerships** that allow states to work together to improve teaching and learning and raise student achievement; and
- serve as a **national clearinghouse** on standards and school reform.

Michael Cohen, President

Matthew Gandall, Executive Vice President

Jean Slattery, Director of Benchmarking

Measuring Up 2005

A Report on Assessment Anchors and Tests
in Reading and Mathematics for

PENNSYLVANIA

ACHIEVE'S
BENCHMARKING INITIATIVE

TABLE OF CONTENTS

Executive Summary	5
Achieve’s Work with Pennsylvania	9
The Achieve Benchmarking Methodology	11
Major Findings	
Review of Pennsylvania’s Assessment Anchors and Eligible Content in Reading	14
Alignment of Assessments to Assessment Anchors and Eligible Content in Reading	30
Review of Pennsylvania’s Assessment Anchors and Eligible Content in Mathematics	36
Alignment of Assessments to Assessment Anchors and Eligible Content in Mathematics	47
Appendix: Biographies	54

EXECUTIVE SUMMARY

Achieve, Inc., is a bipartisan, non-profit organization created by the nation's governors and corporate leaders to help states raise their academic standards, improve their assessments and strengthen accountability to prepare all young people for postsecondary education, work and citizenship. A principal part of Achieve's mission is to provide state policymakers with an independent expert review of the quality of their standards and tests.

Measuring Up 2005: A Report on Assessment Anchors and Tests in Reading and Mathematics for Pennsylvania is the second time Achieve has worked with Pennsylvania. Achieve reviewed the state's academic standards and the alignment of its transitional assessments in summer 1999, when the state was in the process of moving to a standards-based system of assessments, and recommended ways Pennsylvania could strengthen the content and alignment of its new assessments.

The federal No Child Left Behind Act (NCLB) brought new challenges and opportunities for the Pennsylvania System of School Assessment (PSSA). Like many states, Pennsylvania was faced with adding assessments in grades 4, 6 and 7 in reading and mathematics. However, unlike most states, Pennsylvania did not respond to NCLB by simply adding required tests; rather, it decided to take the unusual action of culling core content from its overall academic standards and targeting this content on its assessments. This attempt to focus on the durable, lifelong knowledge and skills that students should acquire and retain is a worthwhile effort.

At the request of state officials, Achieve reviewed Pennsylvania's newly established Assessment Anchors and related Eligible Content statements in reading and mathematics and analyzed their alignment with the state's assessments. This report summarizes Achieve's findings and provides recommendations for strengthening Pennsylvania's comprehensive system of standards and tests.

RESULTS FOR PENNSYLVANIA

- **Pennsylvania has generally identified the most essential content for inclusion in its Assessment Anchors and Eligible Content statements.**

Achieve conducted a series of reviews of Pennsylvania's Assessment Anchors and Eligible Content statements, and the state took full advantage of our recommendations. Each successive version of the state's documents in reading and mathematics was more robust and cohesive than the previous one. For example, in reading, the state persisted until its anchors and related content reflected the more demanding levels of performance described by the larger academic standards. In mathematics, the state sharpened the focus on essential mathematics by reducing redundancy and increasing the precision of the Eligible Content statements.

- **The Assessment Anchors and accompanying Eligible Content statements in reading and mathematics align well with Pennsylvania's overall academic standards.**

In pegging core knowledge and skills for its Assessment Anchors, Pennsylvania was restricted to content contained in its overarching academic standards because it had to

maintain alignment between the new anchors and the existing standards. The state was successful in stipulating and clarifying content for the anchors that was consistent with the content described in the larger academic standards.

- **Pennsylvania’s Assessment Anchors and Eligible Content statements are clear and measurable and encompass a manageable amount of content.**

The Eligible Content statements rightly focus on the results, rather than the process, of learning and usually succeed in bringing the kind of clarity and specificity to the Assessment Anchors necessary for constructing test items able to assess the actual intent of the anchors. The added detail supplied by the content statements also helps keep the “grain size” of the documents uniform.

- **The Assessment Anchors and Eligible Content statements will help teachers develop lessons and classroom assessments.**

Pennsylvania made steady progress in fine-tuning its documents so they would lend themselves to supporting focused, effective teaching and learning. One advantage of the anchors is that they are broken out by grade level — teachers can readily see what students were expected to learn prior to their class and what they are expected to know and be able to do the following year.

By taking Achieve’s suggestion to group topics in mathematics, Pennsylvania is encouraging the design of rich, connected instructional units in lieu of disconnected lessons on isolated topics. Moreover, the state is using its Web site to good advantage, such as by bypassing simplistic “test-prep” tasks in favor of constructing meaningful assignments in reading.

The format highlights pertinent information by linking an Assessment Anchor to related content statements and by cross-mapping the anchor to the text of the relevant academic standard. In mathematics, the state took an important step when it reorganized the content of its academic standards — regrouping 11 standards to form five major strands. These streamlined strands bolster the state’s goal of centering on the durable knowledge and skills students must acquire to be mathematically literate in an information-based, technologically driven society.

- **Pennsylvania should continue to fine-tune the Eligible Content statements so they show a consistent pattern of increasing rigor across grades.**

The state will want to ensure it has constructed a steep enough learning curve so all of its students end up being college and work ready. Lack of careful attention to rigor could undermine the state’s efforts to raise expectations across the board and turn state assessments into minimum competency tests. For example, although some repetition in reading skills is expected — students at all grades need to grasp the gist of text — care must be taken to describe the evolution of skills through the grades to ensure teachers are not settling for surface attention to text and minimal, routine performances. Literary analysis, comprehension of non-fiction text and vocabulary, for example, need to show increasing challenge across grades. While the Eligible Content statements in mathematics taken as a whole describe an adequate progression of knowledge and skills, more could be done to underscore key concepts and adjust the balance of the statements, giving less attention to procedural knowledge and more to conceptual understanding, reasoning and

problem solving. Achieve urges the state to build up its geometry strand across the grades and include content beyond Algebra I and geometry in its grade 11 anchors.

- **Pennsylvania’s tests in reading and mathematics at grades 3, 5, 8 and 11 are strongly aligned to the Assessment Anchors and Eligible Content statements.**

Items on Pennsylvania’s tests in both reading and mathematics are a good match to the content and performances described by the anchors and the related content statements. Achieve found the state has generally succeeded in crafting clear and specific content statements. Reviewers also found the reading passages on the state tests meet the expectations of the anchors in that they include a balance of genres and an increase in complexity across the grades.

- **Pennsylvania made the right decision in including constructed-response items on its tests, but it will need to ensure all are of high quality.**

Well-crafted constructed-response items provide students with the opportunity to demonstrate their ability to respond to complex performances, which typically require advanced knowledge and skills — aspects of the Assessment Anchors that are difficult to assess with multiple-choice questions. As a result, they have a positive influence on classroom instruction, prompting teachers to ask demanding questions of their students. Another positive feature of constructed-response items is that they reflect — more closely than do multiple-choice items — the kind of work expected in college courses and the workplace that puts a premium on students’ ability to analyze, synthesize and evaluate information and apply mathematics. The quality of the state’s constructed-response items varied, and Pennsylvania will want to verify that test developers adhere to state criteria in designing items and related scoring guidelines.

RECOMMENDATIONS FOR MOVING FORWARD

As Pennsylvania continues to build a rigorous and aligned system of standards and assessments, Achieve recommends the state take the following actions:

- ✓ **Strengthen the progression of the Assessment Anchors and Eligible Content statements from one grade to the next so a rigorous trajectory of knowledge and skills is readily apparent.**

Progression in standards is paramount. Pennsylvania should spare no effort in demonstrating that higher grades have more rigorous expectations than lower grades. To delineate a rising demand of expectations across the grades in both reading and mathematics, Achieve recommended the state develop a matrix for its Assessment Anchors and Eligible Content statements that traces each content strand from one grade to the next, spelling out the new knowledge and skills expected at each grade. The state already has begun to act on Achieve’s advice and will soon make these matrices available to teachers.

- ✓ **Provide examples of items and samples of grade-level text to clarify all Eligible Content statements and underscore their increasing rigor across the grades.**

Achieve also recommended that Pennsylvania make the rigor of its grade-level expectations crystal clear by including sample items in reading and mathematics as well as examples of text that students must be able to read and comprehend. Without sample items and reading passages, it is difficult for teachers and parents to grasp the level of performance that the state expects of its students. Pennsylvania has taken Achieve's advice and is already at work preparing an "Item Sampler" and "Item Bank" of released test items.

- ✓ **Develop grade-level standards or high school course standards for missing grades (K, 1, 2, 4, 6, 7, 9, 10 and 12) when Pennsylvania next revises its academic standards.**

Pennsylvania has not revised its academic standards since 1999, and the state will want to take advantage of the next revision cycle to update required content and complete a K–12 continuum. Knowledge and skills in K–2 lay the foundation for future learning, and recent research — unavailable in 1999 — has uncovered optimal ways of delineating learning in the early grades, which Pennsylvania will want to incorporate. At the other end of the spectrum, we now have data that speak to the necessity of students taking rigorous coursework in the core academic areas. It makes sense for the state to develop a complete set of standards to promote excellence and equity across the state, raising the level of proficiency for all students and closing the achievement gap between subgroups and the larger majority.

- ✓ **Ratchet up the cognitive demand of the Assessment Anchors, Eligible Content statements and assessments over time and in concert.**

All states are faced with the challenge of having to prepare their graduates to a much higher level of proficiency in English language arts and mathematics than at any previous time in our nation's history. Our immersion in a global economy, fueled by information and technology, has totally changed and will continue to change the way we live and work. High-growth jobs likely to provide a middle class income already require solid preparation and well-developed skills in core subjects, especially in reading, writing and mathematics. Like all states, Pennsylvania has little choice but to continue steadily raising the rigor of its standards and assessments if it wants to ensure its graduates are able to compete and thrive in this new environment.

ACHIEVE'S WORK WITH PENNSYLVANIA

Achieve was established after the 1996 National Education Summit by the nation's governors and business leaders to provide advice and assistance to state policy leaders on issues of academic standards, assessments and accountability. Under the auspices of Achieve's Benchmarking Initiative, 20 states have sought Achieve's external reviews of state education policy since 1998.

Pennsylvania first sought Achieve's assistance in summer 1999, when the governor and the secretary of education requested that Achieve undertake an evaluation of Pennsylvania's academic standards and the alignment of its transitional assessments to those standards. (The state was in the process of transitioning from its previous testing system to a new standards-based assessment system, and Achieve's review was meant to identify ways in which the alignment of the new assessments to the standards could be strengthened.)

The advent of No Child Left Behind (NCLB) brought new challenges and opportunities for the Pennsylvania System of School Assessment (PSSA). Like many states, Pennsylvania was faced with adding assessments in grades 4, 6 and 7 in reading and mathematics and developing new assessments in science. However, unlike most states, Pennsylvania did not respond to NCLB by simply adding required tests; rather, it decided to take the unusual action of culling out core content from its overall academic standards and targeting this content on its assessments. The resulting grade-by-grade Assessment Anchors and Eligible Content in reading and mathematics (and eventually science) are meant to serve two purposes: (1) identify the foundational knowledge and skills students must acquire to continue to make intellectual headway in these subject areas, and (2) alert teachers to critical content deserving of special emphasis. In other words, by focusing instruction and state-level assessment on essential, durable knowledge, the state hopes to improve the depth of student learning.

The chronology of events in Achieve's work with Pennsylvania provides important context for understanding our findings and recommendations. At the request of then Secretary of Education Vickie Phillips, Achieve undertook a preliminary review of the state's draft documents in late summer 2003 for reading and mathematics to help the state make optimal content selections from its academic standards so expectations for student learning would be more focused but not significantly diminished. Upon receiving Achieve's recommendations, Pennsylvania undertook a thorough review and revision of its Assessment Anchors and related content in both subject areas.

In spring 2004, Achieve bid on and was awarded the state's contract to complete a detailed analysis of Pennsylvania's revised Assessment Anchors and Eligible Content statements in reading and mathematics. The contract also required that Achieve conduct an alignment study of the degree to which Pennsylvania's assessments in grades 3, 5, 8 and 11 measured the content of the revised anchors and eligible content. Achieve conducted its evaluation of the Assessment Anchors and Eligible Content statements in summer 2004 and its alignment review in September 2004. It is helpful in interpreting the findings and recommendations set forth in this report to understand that the state has made a series of painstaking revisions to its Assessments Anchors and Eligible Content statements based on recommendations Achieve provided at each phase of this review process. The state resisted the temptation to set its documents in place prematurely and deserves recognition for viewing them as works in progress, taking advantage of successive opportunities to make thoughtful, incremental improvements.

Achieve was not able to conduct a second comprehensive review of the state's most recent edition of its documents (November 2004); thus, this report centers on our evaluation of the state's summer 2004 edition. However, we have made note of significant steps Pennsylvania has since taken to improve its Assessment Anchors and Eligible Content statements.

THE ACHIEVE BENCHMARKING METHODOLOGY

THE EVALUATION OF PENNSYLVANIA’S ASSESSMENT ANCHORS AND ELIGIBLE CONTENT

It is important to understand that this particular analysis is not the comprehensive review of Pennsylvania’s academic standards as compared with benchmark documents that Achieve typically undertakes. Rather, it is an evaluation of the state’s effort to cull a subset of essential, foundational core knowledge and skills in reading and mathematics and to target these for instructional emphasis and assessment. In selecting areas of focus for its Assessment Anchors and Eligible Content, the state was constrained by the content delineated in its current academic standards: Its Assessment Anchors and Eligible Content statements must be aligned to these current academic standards. Achieve, therefore, concluded it would be most helpful to the state if our reviewers made use of Achieve’s benchmark documents in two complementary ways:¹ first, to help determine if Pennsylvania captured the most important content from its academic standards for inclusion in its Assessment Anchors and Eligible Content; second, to make note of significant content contained in Achieve’s benchmarks, but not in Pennsylvania’s academic standards, for the state to consider for inclusion when it next revises its standards.

To help reviewers conduct a systematic analysis and to ensure Pennsylvania’s concerns were addressed, Achieve formulated a set of guiding questions that directed the attention of our reviewers to evaluating Pennsylvania’s Assessment Anchors on the basis of the following criteria:

- **Focus:** Are these the most important measurable outcomes?
- **Overall Organization and Format:** Are these helpful to the reader?
- **Vertical Alignment:** Do knowledge and skills appropriately progress across grade levels?
- **Clarity:** Are the Assessment Anchors easy to read and understand?
- **Measurability:** Are the Assessment Anchors appropriate for large-scale testing?
- **Manageability:** Can the Assessment Anchors be taught and learned by the April test administration?
- **Coherence:** Do the anchors and accompanying materials help educators understand how to make the connection among standards, assessment, curriculum and instruction?

However, in the end, guiding questions are only a tool; the quality of Achieve’s reviews rests on the expertise of its reviewers, all of whom have deep experience in evaluating standards and assessments in their subject area. (Brief biographical sketches of our reviewers are included as an appendix to this report.)

¹ **Achieve’s benchmark standards** in English language arts are those from California (1997) and Massachusetts (2001) and also include early literacy standards from North Carolina (1999), Texas (2001) and New Standards (1999). In mathematics, Achieve’s benchmarks are those from Indiana (2000) and Singapore (2001), as well as its own document *Foundations for Success* (2002), which details the mathematics that we believe all students should be expected to know before leaving 8th grade. Under the auspices of the American Diploma Project, Achieve produced college and workplace readiness benchmarks for English and mathematics.

THE ALIGNMENT STUDY

Alignment is a measure of the extent to which standards and assessments agree and the degree to which they work in conjunction to guide and support student learning. It is not a question that yields a “yes” or “no” response; rather, alignment is a considered judgment based on a number of factors that collectively determine the degree of match between a state’s standards and the assessment used to gauge if students are meeting those standards. At its core, Achieve’s analysis answers two key questions for Pennsylvania: (1) Can everything on the assessments be found in the state’s Assessment Anchors and Eligible Content statements? (2) In addition (and conversely), do the assessments do an effective job of measuring the knowledge and skills set forth in the anchors and related content statements?

To determine how closely each Pennsylvania assessment was aligned to the related grade-level Assessment Anchors and Eligible Content, Achieve convened two teams of content experts who followed a subject-specific, stepwise procedure (protocol) that Achieve has used to evaluate numerous assessments in more than a dozen states.

In the first phase of the review process, a team of content experts evaluates each individual test item to determine (1) if it actually measures the indicator to which it was assigned by the test developer, (2) how well it matches the content and performance described in the related standard, (3) whether it is fairly constructed and (4) how intellectually challenging it is. These are key issues. The information gained from a test is no better than the collection of items that make it up. If an item measures content and skills beyond what is contained in the standards, it is far less likely that it will have been taught in classrooms. Similarly, an item that is flawed for reasons such as having no right answer, more than one right answer, a misleading graphic or implausible distracters will not give accurate information about students’ performance. Tracking the level of cognitive demand that each individual item poses also is critical. If a test is truly standards based, it should have a mix of basic and more challenging items that reflect the range of concepts and skills spelled out in the standards so differences in the performance of proficient and non-proficient students can be detected. In summary, Achieve’s item-by-item analysis not only yields valuable information about critical aspects of alignment but also provides quantitative data that contribute to the judgments made with respect to the overall balance and rigor of a test, as described below.

In the second phase of the alignment study, content experts take a more holistic view of the test to judge if it is balanced overall and if it is appropriately rigorous for the grade level. Moving away from the item level, reviewers consider the test one major strand — such as Literary Analysis or Geometry — at a time and appraise the collection or set of items that are meant to assess each strand.

To judge how balanced the set of items mapped to each standard is, experts ask, “Does this set of items succeed in measuring the breadth and depth of content and skills described in the strand?” In other words, “To what extent does the set of items assess the key content and skills in the standard?” Because a single on-demand test cannot assess all the Eligible Content statements that make up the state’s Assessment Anchors, it is crucial to determine how well the items on a test sample the most essential indicators. Content experts also examine the reading passages as a set to ensure that varied types of literary and informational text forms are represented. Where the state includes a direct assessment of writing on its tests, experts check to see that the range of writing prompts

reflects the variety of genres represented in the standards. In evaluating the rigor of a test, content experts follow the same general procedure they use when evaluating balance. They compare the overall intellectual demand encompassed by a set of items with the level of intellectual demand described in the related standard. Looking at each standard in turn, they ask, “Does doing well on the item set, which measures this standard, mean the student has mastered the challenging material contained in the standard?” Because experts rated each item earlier in the process as to its level of cognitive demand, they can determine if an item set has a span of difficulty appropriate for the grade level. Content experts also review the reading passages as a set to determine if they have a span of demand appropriate to the grade level tested, and when applicable, they review writing prompts to ensure the genre, topic and characteristics of a response that will meet standards are clearly communicated in the directions to students.

At the close of the analysis, reviewers look across the standards and consider the test as a whole to determine how well it measures the knowledge and skills described by the standards and how rigorous the test is overall.

THE CONTENT OF THIS REPORT

In writing this report, Achieve synthesized reviewers’ findings regarding Pennsylvania’s Assessment Anchors and Eligible Content statements and the alignment of the state’s assessments to the anchors and eligible content. These studies were conducted by teams of national experts with significant experience in analyzing academic standards and tests. The findings in this report represent consensus opinions of Achieve’s experts, but final judgments and conclusions rest with Achieve. In addition to this summary report, Achieve has prepared a detailed technical report for the Pennsylvania Department of Education. Because the technical report contains references to secure test items, it is confidential.

MAJOR FINDINGS: REVIEW OF PENNSYLVANIA’S ASSESSMENT ANCHORS AND ELIGIBLE CONTENT IN READING

Achieve reviewed Pennsylvania’s initial draft of its reading document in fall 2003 and recommended the state make numerous changes. We found in our subsequent summer 2004 review that the state had succeeded in identifying the most important content for each grade, making the language of the document clearer and more precise, and ensuring that content statements are measurable in a large-scale testing situation. We also concluded, however, that some additional changes would further enhance the state’s articulation of its Assessment Anchors and Eligible Content statements.

The state continued to make revisions in response to Achieve’s summer 2004 review. Although we were unable to conduct a second comprehensive analysis, in compiling our detailed technical report (November 2004), we tried to acknowledge significant changes the state had made up to that time.

In the main, this report summarizes the results of Achieve’s summer 2004 review of the Pennsylvania Assessment Anchors and Eligible Content statements. However, we have made a concerted effort in this report to acknowledge subsequent revisions that have upgraded the quality of the state’s reading document.

We found, overall, that the changes made to the content, organization and language of the Assessment Anchors in the November 2004 edition, as currently shown on the state’s Web site, have made the statements clearer and the content deeper. In fact, with each iteration, Pennsylvania’s reading document has become more precise and robust.

Based on our benchmark standards, Achieve also has made several recommendations for the inclusion of significant content in the Assessment Anchors that currently is not contained in the state’s academic standards. These suggestions will be important for the state to consider when it next revises its academic standards — they are not intended for this edition because Achieve recognizes the importance of maintaining alignment between the state’s anchors and its current academic standards document.

STRENGTHS OF THE ASSESSMENT ANCHORS

- **The Assessment Anchors in reading align well with Pennsylvania’s academic standards.**

Pennsylvania had the challenge of writing clear Assessment Anchors and related Eligible Content statements that would capture and clarify essential core content while staying true to the existing academic standards adopted in 1999. Pennsylvania has done a good job of constructing Assessment Anchors that are precise and clearly represent its overarching academic standards in reading. In earlier versions of the Assessment Anchors, Achieve expressed some concern that the state had not always reached the higher levels of performance intended by the reading standards. The state’s latest revisions better reflect the levels of performance called for by the standards.

- **Overall, Pennsylvania has identified the most important content for inclusion in its Assessment Anchors and Eligible Content at each grade.**

In general, the content of the grade-level Assessment Anchors comprises the key knowledge and skills, such as “comprehension and analysis of both fiction and non-fiction text” and “literal and inferential understandings and vocabulary skills,” over which students must gain increasing control. The anchors also succeed in identifying those elements of reading, both fiction and non-fiction, that are best measured in a large-scale mode. The state responded to Achieve’s recommendations by adding essential content as shown in the table below:

Content Recommendations Already Enacted by Pennsylvania

ACHIEVE RECOMMENDATION (August 2004)	PENNSYLVANIA RESPONSE (November 2004)
Adjust content from the first three academic standards, including: <ul style="list-style-type: none"> • Citing examples of support from text 	Added a related statement of eligible content. For example, at grade 5, statement R5.A.2.3.2 now reads: <i>Cite evidence from text to support generalizations.</i>
<ul style="list-style-type: none"> • Identifying and evaluating essential vs. non-essential information (add to grade 5 and up) 	<p>Added Assessment Anchors on essential vs. non-essential information to its higher-grade-level anchors. For example, at grade 5, R5.B.3.2 reads: <i>Distinguish between essential and non-essential information within or across text.</i> The corresponding statement of eligible content is R5.B.3.2.2: <i>Identify stereotypes where present.</i></p> <p>Also added topics to help ensure students gain necessary skills in analyzing some of the techniques that authors use to develop their arguments.</p>
<ul style="list-style-type: none"> • Identifying main idea (grade 3) 	Added main idea to grade 3. Eligible Content statement R3.A.1.4.1 now reads: <i>Identify stated or inferred main ideas and relevant supporting details from the text.</i>
<ul style="list-style-type: none"> • Revising vocabulary (remove lower-level skills and add idioms and figurative language to grades 8–11) 	<p>Revised the content of its vocabulary statements, removing, for example, the skill of defining compound words from its vocabulary statements.</p> <p>Also revised its statement on figurative language at grade 8 to more clearly read: <i>Identify the author’s purpose for and effectiveness at using figurative language.</i></p>

- **The Assessment Anchors are formatted in such a way as to be easily understood and used by Pennsylvania educators.**

At each grade level, the Assessment Anchors on which this report is based are organized under two broad reporting categories and related subcategories, which vary by grade level:

- A. Comprehension and Reading Skills
- B. Interpretation and Analysis of Fiction and Non-fiction Text

Each Assessment Anchor is followed by a series of related Eligible Content statements that describe the knowledge and skills on which the state will base its assessments.

The format of Pennsylvania's reading document is easy to understand and use. It is helpful to have each page include pertinent information such as the Assessment Anchor, the map to the relevant Pennsylvania standard(s) and the more specific Eligible Content statements.

Achieve had recommended that the state also include the text of the standard (not just the number) for reference. The state has now done so, and the inclusion of the standard's text makes the cross-mapping with the Assessment Anchors readily apparent.

- **The Assessment Anchors present concepts and skills appropriate for testing in a large-scale situation.**

Unlike the state's 2003 draft — in which there were some skills, strategies and content that would be difficult to assess on a large-scale assessment — the reading Assessment Anchors now are written in such a way as to be easily measurable in a multiple-choice and constructed-response format on a large-scale assessment. The state has clearly thought through each anchor to ensure its measurability.

In revising its 2003 preliminary draft of the anchors, the state eliminated standards that cannot be well assessed on a large-scale, on-demand test. For example, those standards addressing research skills have been dropped, as they are better addressed at the classroom level. The decision to retain only those skills that are most effectively assessed at the state level is wise and respects the fact that not all of the most important areas in reading are amenable to large-scale assessment — not all that is important to teach and learn can be tested in a paper-and-pencil format.

- **Pennsylvania's Assessment Anchors and Eligible Content statements are written in easy-to-understand, specific language that clearly conveys expectations for students.**

Constructing Assessment Anchors and Eligible Content statements that clearly convey the content that is "fair game" for assessment, employ a range of action verbs to capture the variety of performances expected of students, and achieve an appropriate and consistent "grain size" is a difficult but essential task.

Often, achieving clarity is a matter of increasing specificity. Achieve recommended, for example, that the state review sections of the reading document to be sure the Eligible Content statements associated with each Assessment Anchor are inclusive and not simply exemplary. For example, in R5.A.2.7, the Assessment Anchor read: *Analyze text organization including sequence, comparison/contrast, cause & effect, problem/solution, the headings, graphics and charts to derive meaning*. Yet, the Eligible Content (R5.A.2.7.1) only read: *Items may address information found in a text subsection, including graphics and charts*. Pennsylvania responded by replacing R5.A.2.7.1 with the following set of statements that include all the concepts addressed in the Assessment Anchor:

-
- R5.B.3.3** *Identify text organization, including sequence, question/answer, comparison/contrast, cause/effect or problem/solution.*
- R5.B.3.3.2** *Use headings to locate information in a passage, or identify content that would best fit in a specific section of text.*
- R5.B.3.3.3** *Interpret graphics and charts, and make connections between text and the content of graphics and charts.*

Achieve also found that strong and focused performance words (action verbs) were not consistently used to begin each statement. Pennsylvania has since revised its statements of eligible content so they all follow a parallel structure and begin with clear action verbs that state the performance expected of students on the state's tests.

The state also clarified expectations by replacing the separate section originally devoted to author's purpose and instead included that concept within both the fiction and non-fiction sections. This revision had the effect of emphasizing the point that students need to be mindful of the author's purpose when they read both fiction and non-fiction texts.

As a result of these revisions, the anchors and related content statements are mostly clear and specific — neither too broad nor too narrow — and should provide guidance to teachers in making the instructional decisions necessary for daily lesson planning and classroom-level assessment.

- **The Assessment Anchors in reading at each grade level present a manageable amount of content, skills and implied strategies.**

The Assessment Anchors do a good job of including a reasonable and manageable amount of knowledge, skills and strategies for educators to teach by the time the test is administered in early spring. If teachers view each Assessment Anchor as a separate and isolated topic for instruction, they may feel that there are too many to fully address by April. However, all of the Assessment Anchors are intended to be taught in concert — effective instruction will combine many of these skills and strategies. Moreover, due to the nature of the reading process, many of the Assessment Anchors repeat from grade to grade. Teachers are not responsible for teaching these topics anew each year, but rather they are building on what previous years' teachers have taught by working with increasingly challenging texts.

- **The Assessment Anchors, Eligible Content statements and accompanying materials help educators understand how to make the connection among standards, assessments, curriculum and instruction.**

The Assessment Anchors do a good job of making the connection between standards and assessment. The state's PowerPoint presentation that is currently on its Web site ("Assessment Anchors: Get Ready, Set, Go!") is evidence of the state's effort to connect strong instruction with assessment by focusing on the construction of meaningful assignments in reading that are not simplistic "test-prep" materials. The continuation of such support to districts and schools will help avoid the major pitfall of narrowing the scope of assessment, thus narrowing instruction. Pennsylvania should extend its outreach by presenting some alternative ways for schools to assess and monitor the instruction and attainment of those standards that are not amenable to large-scale assessment, such as research skills.

The state also may want to consider whether it should develop related curriculum materials to help teachers envision and implement the kind of practices that have been demonstrated to be effective in helping students master essential skills and strategies. An additional benefit of preparing curricular materials would be to weave in the concepts of language, speaking, listening, research and experience with non-print media that are not assessed on the reading and writing portions of the PSSA, thereby helping ensure that these are not overlooked in instruction. Achieve's American Diploma Project underscored the importance of students developing proficiency in these areas as prerequisites for their success in continuing education and in today's information-based workplace.

AREAS FOR IMPROVEMENT

- **Pennsylvania has strengthened the progression of its Assessment Anchors and Eligible Content statements but has not gone far enough to exhibit a consistent pattern of increasing levels of cognitive demand across the grade levels.**

A carefully wrought progression of knowledge and skills across the grades is essential to ensure K–12 standards are sufficiently rigorous to prepare students for the demands of college-level work and employment in today's knowledge-based workplace. Lack of careful attention to rigor is likely to turn state assessments into minimum competency tests, thereby defeating the goal of standards-based assessments meant to raise expectations for all students in all grades.

In some areas, repetition is expected in English language arts. We expect students to answer literal questions about the details of texts, make inferences, draw conclusions and identify the author's purpose at all grade levels. Even at higher grades, we want to ascertain whether students grasp the literal meaning of text because we expect them to be reading and comprehending texts of greater complexity. However, some progression of skills is still expected.

In general, there are at least three ways to increase the cognitive demand of reading skills through the grades.

1. Increase the amount and complexity of the content (from simile to irony, for example).
2. Increase the demand of the performance (from simple identification to explanation to interpretation to analysis to evaluation, for example).
3. Increase the complexity of the reading passages.

Without such a progression in assessment expectations, teachers may settle for surface attention to texts and minimal performance expectations. In the revision process in which Pennsylvania has been engaged over the past year, the state has succeeded in crafting many anchors that delineate incremental cognitive demand in content — for example, the increasing addition of suffixes and prefixes at R.3–8 and 11.A.1.2.1 and the increasing list of elements in the areas of character, plot, setting, theme, tone and style at R.3–8 and 11.B.1.1.1. In its revised Assessment Anchors for identifying various components of fiction and non-fiction texts, the state shows a good progression of expectations by carefully choosing performance verbs, as the table on the next page demonstrates. (Note that italics have been added to highlight the progression of the performance verbs.)

Progression of Expectation for Literary Analysis

R3.B.1.1.1	R4.B.1.1.1	R5.B.1.1.1	R6.B.1.1.1	R7.B.1.1.1	R8.B.1.1.1	R11.B.1.1.1
<i>Identify the following in fiction and literary non-fiction texts:</i>	<i>Identify the following in fiction and literary non-fiction texts:</i>	<i>Identify and compare the relationships among the following within or across fiction and literary non-fiction texts:</i>	<i>Identify and compare the relationships among the following within or across fiction and literary non-fiction texts:</i>	<i>Describe and interpret the relationships among the following within or across fiction and literary non-fiction texts:</i>	<i>Describe and interpret the relationships among the following within or across fiction and literary non-fiction texts:</i>	<i>Describe, analyze and evaluate the relationships among the following within or across fiction and literary non-fiction texts:</i>

Often, however, Pennsylvania's Assessment Anchors as they are currently drafted (November 2004) retain relatively the same level of cognitive demand through the grade levels. If the Assessment Anchors do not specifically define levels of challenge and complexity, then the result is only a loose definition of what should be discussed and practiced in classrooms. Issues of level of cognitive demand, complexity, length and quality of an acceptable response are then left to the individual teacher, who will make his or her own interpretation that may or may not match the state's intent. If the levels of demand and complexity do not appear to develop within the Assessment Anchor statements and the Eligible Content statements, the state will want to make sure that it sufficiently articulates the expected increase in challenge across grade levels by providing high-quality example items that document an increasing level of challenge and sample reading passages (or descriptions of reading passage levels) that show the expected increase in reading level.

Progression in standards is paramount. Pennsylvania should spare no effort in demonstrating that higher grades have more rigorous expectations than lower grades.

To consider progression in the earlier drafts of the Pennsylvania Assessment Anchors, Achieve reviewers created a table (or matrix) to show how similar topics progressed from grade 3 through grade 11. Achieve recommended that the state build on Achieve's model and lay out, side by side, common content-area statements that run across the grade levels to ensure that the language remains consistent when there is no intended change of meaning and, conversely, that the language changes when the state wants to show an increase in the level of challenge from grade to grade. This kind of matrix — or continuum of knowledge and skills — has the potential to be a useful tool for the state to use with curriculum planners and educators involved in instruction across grade levels. Pennsylvania has decided to act on Achieve's advice and will make the matrix available to teachers as soon as possible.

To assist Pennsylvania in strengthening the vertical alignment of its Assessment Anchors, Achieve made specific recommendations in its summer 2004 technical report aimed at clarifying and ensuring the appropriateness of progression in three key areas: literary analysis, text comprehension (non-fiction) and vocabulary.

Literary Analysis

Achieve had noted that Pennsylvania’s treatment of literary analysis would be enhanced if there were more clarity in the progression of the cognitive demand of content across grade levels. Reviewers, for example, argued that while students at all grades should consider plot, character and setting, the expectations could be made more demanding at the middle school and high school levels without sacrificing a match to the existing standards (being that these are fairly general and inclusive in their language on literary elements). They suggested the following changes:

- a. Make students at grade 7 and above aware of the function of subplots and secondary characters;
- b. Include the function of setting in creating mood at the higher grades; and
- c. Call for students in high school to make judgments about the static or dynamic nature of the characters, their plausibility and the complexity of their development — sending a clear signal that secondary students need to be able to evaluate texts as well as comprehend them.

The state accepted these suggestions and improved the vertical articulation of significant content in this strand.

Achieve also recommended two other significant changes to clarify content progression. We urged the state to introduce concepts consistently across Assessment Anchors or reporting categories (for example, “theme” at grade 3 and grade 6). Pennsylvania corrected the inconsistency and now introduces theme for the first time at grade 5, in terms of both identifying the theme in a summary of the text and identifying it as part of the literary analysis statements.

In addition, we recommended the state be on guard for instances of “artificial” progression. Sometimes the state attempted to show a progression from lower- to higher-level skills, but the progression was not meaningful as worded. For example, the grade 4 Assessment Anchor:

R4.A.1.3.1: *Make inferences and draw conclusions based on explicit information from the text.*

became at grade 5 ...

R5.A.1.3.1: *Make inferences and draw conclusions based on explicit and/or implicit information from the text(s).*

The addition of “implicit” from grade 4 to grade 5 was not particularly meaningful because inferences always require a leap from the explicit text to what is implicit in or implied by the text. Pennsylvania concurred with Achieve’s recommendation and removed the references to implicit and explicit inferences from texts at all grades.

Text Comprehension: Non-fiction

The section of the Assessment Anchors that dealt with the comprehension of non-fiction text (A.2.3 and A.2.4 in previous versions of the Assessment Anchors) did not convey a trajectory of increasing knowledge and skills, especially with respect to the development

of main idea. Main idea still lacks a progression in terms of the action or content of the Eligible Content statements, as shown below:

Grade 3: R3.A.2.4.1: *Identify stated or implied main ideas and relevant supporting details from text.*

Grade 4: R4.A.2.4.1: *Identify stated or implied main ideas and relevant supporting details from text. (Same as previous grade)*

Grade 5: R5.A.2.4.1: *Identify and/or interpret stated or implied main ideas and relevant supporting details from text.*

Grade 6: R6.A.2.4.1: *Identify and/or interpret stated or implied main ideas and relevant supporting details from text. (Same as previous grade)*

Grade 7: R7.A.2.4.1: *Identify and/or interpret stated or implied main ideas and relevant supporting details from text. (Same as previous grade)*

Grade 8: R8.A.2.4.1: *Identify and/or interpret stated or implied main ideas and relevant supporting details from text. (Same as previous grade)*

Grade 11: R11.A.2.4.1: *Identify and/or interpret stated or implied main ideas and relevant supporting details from text. (Same as previous grade)*

Achieve has recommended that the state may want to consider whether it is important to repeat Eligible Content statements across multiple grade levels when there is no change in content. Some states specify that at each grade level students are expected to show continued mastery of the previous grades' standards. For topics such as identifying the main idea of text, the state may decide that it is important enough to include at all grade levels. But a clearer progression could still be shown. Perhaps, for example, it could be assumed that students at the highest grade levels would not need to identify stated main ideas and the Eligible Content statements could be revised accordingly to show a progression.

Similarly, the Eligible Content statements related to informational text organization and the use of headers and graphics remain almost identical between grades 3 and 8 as follows:

R3.B.3.3.1: *Identify text organization, including sequence, question/answer, comparison/contrast, cause/effect or problem/solution.*

R3.B.3.3.2: *Use headings to locate information in a passage, or identify content that would best fit in a specific section of text.*

R3.B.3.3.3: *Interpret graphics and charts, and make connections between text and the content of graphics and charts.*

The state may wish to consider ways it can better show the increasing challenge that it expects students to meet across grades 3 through 8.

Despite the lack of progression evidenced in some aspects of non-fiction reading, however, Pennsylvania's November 2004 edition of its reading document demonstrates the state has made progress in articulating a progression of non-fiction reading skills by enriching its section on non-fiction with additional non-fiction comprehension elements (beyond main idea, details and inferences). For example:

Grade 3: R3.B.3.2.1: *Identify exaggeration where present.*

Grade 4: R4.B.3.2.1: (same)

Grade 5: R5.B.3.2.2: *Identify stereotypes where present.*

Grade 6: R6.B.3.2.2: (same)
Grade 7: R7.B.3.2.1: *Identify bias and propaganda techniques where present.*
Grade 8: R8.B.3.2.1: (same)
Grade 11: R11.B.3.2.1: (same as above)
Grade 11: R11.B.3.2.2: *Analyze the effectiveness of bias and propaganda techniques where present.*

Vocabulary

Pennsylvania has responded to some of Achieve’s concerns about repetition and progression in its vocabulary Assessment Anchors. In earlier versions of the Assessment Anchors, grades 6 through 11 included statements on compound words as follows:

R6.A.1.1.3: *Identify meanings of compound words and possessives.*
R7.A.1.1.3: *Identify meanings of compound words and possessives.*
R8.A.1.1.3: *Identify meanings of compound words and possessives.*
R11.A.1.1.3: *Identify meanings of compound words and possessives.*

Achieve suggested the state reconsider the inclusion of this topic, as it seemed insufficiently demanding for grades 8 and 11. As a result of this feedback, the state reconsidered its vocabulary statements and removed those on compound words and possessives.

However, the state decided not to follow Achieve’s recommendation and still includes multiple-meaning words at the higher grades. Achieve reviewers agreed that the identification of a multiple-meaning word is important for early readers (knowing the difference between the bank of a river and the bank that holds money, for example). But reviewers argued that at the middle and secondary grades a more appropriate challenge would be to determine the connotation and denotation of a word in a text or the effectiveness of the use of idioms and colloquialisms. Instead, the state decided to continue to include statements on multiple-meaning words at every grade level (as shown below).

It is still Achieve’s position that unnecessary repetition undermines progression. Currently, the content of Pennsylvania’s vocabulary Eligible Content statements repeats as follows:

Grade 3: R3.A.1.1.1: *Identify meaning of multiple-meaning words used in text.*
Grade 4: R4.A.1.1.1: (same as above)
Grade 5: R5.A.1.1.1: (same as above)
Grade 6: R6.A.1.1.1: (same as above)
Grade 7: R7.A.1.1.1: (same as above)
Grade 8: R8.A.1.1.1: (same as above)
Grade 11: R11.A.1.1.1: (same as above)

Grade 3: R3.A.1.1.2: *Identify a synonym or antonym of a word used in text.*
Grade 4: R4.A.1.1.2: (same as above)
Grade 5: R5.A.1.1.2: (same as above)
Grade 6: R6.A.1.1.2: (same as above)
Grade 7: R7.A.1.1.2: (same as above)
Grade 8: R8.A.1.1.2: (same as above)
Grade 11: R11.A.1.1.2: (same as above)

The state may wish to further consider this vertical repetition of the same content and performance. As stated earlier, the creation of a matrix — in which content and performance that do not change and content and performance that do change show a progression in cognitive demand — would be a helpful tool to the state in considering progression.

- **Pennsylvania’s November 2004 edition of its reading document would be further improved if the state were to tighten overall organization, clarity and grain size.**

Organization

As a final note, it is important to reiterate that Achieve did not conduct a complete review of the state’s November 2004 Assessment Anchors. Our abbreviated review, however, suggests more work should be done in making the organization tighter and more coherent. In the version of the Assessment Anchors that Achieve reviewed in summer 2004, Pennsylvania separated fiction and non-fiction in some sections and grouped them in others. This resulted in the repetition of some concepts (such as vocabulary acquisition, main idea and details, and so on).

Currently, the state has in place the organization that is shown below. (Note that the Assessment Anchors vary somewhat as the grades progress, but the overall organizational structure and areas of content remain the same. The following example is grade 3.)

Organizational Structure:

- i. Reporting category
- ii. Assessment anchor and description
- iii. Eligible content

Reporting Category A: Comprehension and Reading Skills

R3.A.1: *Understand fiction text appropriate to grade level.*

R3.A.2: *Understand non-fiction text appropriate to grade level.*

Reporting Category B: Interpretation and Analysis of Fiction and Non-fiction Text

R3.B.1: *Identify components within text.*

R3.B.2: *Identify literary devices.*

R3.B.3: *Identify concepts and organization of non-fiction text.*

In terms of overall organization, Achieve found that the Eligible Content statements in R3.A.1 are almost identical to the Eligible Content statements in R3.A.2, the only difference being that Anchor 1 addresses fiction and Anchor 2 non-fiction.

The goal of providing separate sections for fiction and non-fiction in Reporting Category A may be to stress the importance of instruction in non-fiction. The state may want to consider, however, whether repeating the same Eligible Content twice is the best approach. Would it better highlight the importance of instruction in non-fiction if the separate section on non-fiction reading emphasized those characteristics that are exclusive to non-fiction reading (such as patterns of organization; text features such as headers, development of arguments; etc.)? The November 2004 edition of the reading document shows that Reporting Category B now includes many of the skills and strategies related to non-fiction and informational-text reading that are specific to reading these kinds of texts and, therefore, could be presented in a separate category. In such a

schema, Reporting Category A could be the general comprehension skills and strategies (such as understanding vocabulary, identifying main ideas and details, making inferences) that are relevant to both fiction and non-fiction (or literary and informational) texts. Reporting Categories B and C could then focus on literary and informational texts respectively.

An organization such as the one shown below, as an example, would accomplish the goals of emphasizing the importance of non-fiction or informational texts and articulating why/how instruction in informational-text reading is different from literary-text reading.

A. Comprehension and General Reading Skills

- A.1 Vocabulary and Word Skills
- A.2 Genres
- A.3 Author's Style and Purpose
- A.4 Topics, Main Ideas and Supporting Details
- A.5 Inference and Conclusions

B. Interpretation and Analysis of Non-fiction

- B.1 Non-fiction Organization and Text Features
- B.2 Non-fiction Concepts (e.g., fact/opinion, ideas to support arguments, etc.)

C. Interpretation and Analysis of Literature

- C.1 Literary Elements (e.g., character, plot, setting and theme)
- C.2 Literary Devices (e.g., figurative language, metaphor, alliteration)

Even if Pennsylvania decides not to consider a new organizing structure, the state may want to make Reporting Category B parallel with Reporting Category A by separating fiction and non-fiction into two separate sections there as well. This change would not require a significant restructuring of the Assessment Anchors, just a re-ordering and renumbering of Assessment Anchors in Reporting Category B. For example, at grade 3, the newly organized Assessment Anchors would be structured as follows:

R3.A: Comprehension and Reading Skills

R3.A.1: *Understand fiction text appropriate to grade level.*

R3.A.2: *Understand non-fiction text appropriate to grade level.*

R3.B: Interpretation and Analysis of Fiction and Non-fiction Text

R3.B.1: *Interpret and analyze fiction and literary texts.*

R3.B.2: *Interpret and analyze non-fiction texts.*

Clarity

Although Achieve was not able to conduct a complete review of the November 2004 reading document, we did find that some of the newly added statements of eligible content did not give a good picture of how they might be assessed on the statewide assessment. Editors should read and try to paraphrase each statement and generate examples of how it might be assessed. Different readers can compare their restatements and examples and rewrite the content statements as needed to help ensure consistency in how they are interpreted by readers.

For example, Eligible Content statement B.1.1.1 begins: *Describe, analyze and evaluate the relationships within or across fiction and literary non-fiction texts ...*. This suggests that there will be no items that address character only or setting only (or any of the other text “components”). Rather, all items will get at relationships between literary elements or text components, and many items will be cross-text items. If this is not the case, then this should be reworded to allow for the possibility of items that ask students to analyze characters, setting or other components on their own and not in relationship to other text elements.

As a further example of wording in which clarification may be needed, some of the subtopics under B.1.1.1 at grade 11 include:

Content:

- Analyze the relationship between content and other components of text (may analyze content across fiction or non-fiction texts).

Topic:

- Analyze topics or subtopics within or across texts to determine a relationship.
- Analyze the relationship between the topic of non-fiction text and components of text (may analyze a topic across fiction or non-fiction texts).

It is not entirely clear what it might look like if a student is analyzing the relationship between content or topic and other text components. What components of the text might be considered — only those listed in B.1.1.1? Also, there is an emphasis in grade 11, R11.B.1.1.1, on comparing relationships across texts. Will paired texts be emphasized in future assessments? An example would be helpful here.

Achieve raised other concerns about clarity in the November 2004 edition. For example, at grade 7, the Eligible Content statement R7.A.1.6.2 reads: *Identify examples of text that support its narrative or poetic purpose*. Achieve reviewers were unsure what the intention of this statement was. Did this mean that students would identify examples from text that support the author’s purpose for writing? Or would students identify examples from text that support their identification of the text as narrative or poetic? Or would students be given several different texts and have to identify one that was written to be, for example, a poem? In response to these concerns, the state has indicated its plans to revise this statement.

Grain Size

As noted previously, Pennsylvania made changes to address Achieve’s earlier concerns with the parallel grain size of Assessment Anchors.

Some of the November 2004 revisions, however, created a grain-size issue within the current Assessment Anchors. With the inclusion of both fiction and non-fiction elements into B.1.1.1, the size of that Assessment Anchor has become somewhat unwieldy and encompasses much more (particularly at the higher grades) than do other Assessment Anchors. If an item is mapped to B.1.1.1, that assignment will not provide much insight into what kind of item it is (being that it could be one of many contents and performances expected). In addition, readers may have trouble focusing on all of the content included in B.1.1.1 because it is not as clearly and specifically organized as other Eligible Content

statements. The state may want to consider numbering these items separately, as it has in the other Assessment Anchors with the other statements of Eligible Content.

To better show this difference in size, it may be useful to show some examples. For example, at grade 11, the first Eligible Content statement addresses a relatively narrow type of vocabulary knowledge:

R11.A.1.1.1: *Identify meaning of a multiple meaning word in text.*

By contrast, statement B.1.1.1 at grade 11 encompasses much, much more:

Describe, analyze and evaluate the relationships among the following within or across fiction and literary nonfiction texts.

Character (also may be called Narrator, Speaker or Subject of a biography):

- Analyze the relationships between characters and other components of the text.
- Analyze character actions, motives, dialogue, emotions/feelings, traits and relationships among characters within or across texts (may analyze characters across fiction and non-fiction texts).

Setting:

- Analyze the relationship between setting and other components of the text.
- Analyze settings across texts (may analyze setting across fiction and non-fiction texts).

Plot (also may be called Action):

- Analyze the relationship between elements of the plot (conflict, rising action, climax/turning point, resolution) and other components of the text.
- Analyze elements of the plot across texts (may analyze plot/action across fiction and non-fiction texts).

Theme:

- Analyze the relationship between the theme and other components of the text.
- Analyze themes across texts (may analyze theme across fiction and non-fiction texts).

Topic (of literary non-fiction):

- Analyze the relationship between the topic and other components of the text.
- Analyze topics across non-fiction texts.

Tone and Style:

- Analyze the relationship between the tone and/or style and other components of the text.
- Analyze the relationship between tone and/or style across texts (may analyze tone/style across fiction and non-fiction texts).
- Describe, analyze and evaluate the relationships among the following within or across non-fiction texts.

Content:

- Analyze differing viewpoints within non-fiction text.
- Analyze the relationship between content and other components of text (may analyze content across fiction or non-fiction texts).

Topic:

- Analyze topics or subtopics within or across texts to determine a relationship.
- Analyze the relationship between the topic of non-fiction text and components of text (may analyze a topic across fiction or non-fiction texts).

Style/Tone:

- Analyze the effectiveness of the author's use of style and tone within non-fiction text.
- Analyze the relationship between author's use of style and tone and other components of text (may analyze style and tone across fiction or non-fiction texts).

Arriving at a relatively consistent grain size is advisable because it supports clarity.

SUMMARY OF RECOMMENDATIONS

- ✓ **Construct a matrix that charts the progression of Assessment Anchors and Eligible Content statements so that the development of knowledge and skills contained in each strand can be readily traced from one grade to the next.**

To fine-tune the next edition of the Pennsylvania's Assessment Anchors and Eligible Content statements, Achieve recommends that the state create a cross-grade matrix that traces each content strand through the grades, indicating what new knowledge and abilities are expected at each grade. Several states, including Massachusetts and Ohio, have included matrices in their standards documents to delineate the progression across the grades. Maryland has used a similar approach in laying out its Voluntary State Curriculum.

Achieve sees a number of advantages to this approach because a matrix, by its nature, directs attention to sequencing and specificity.

Having a mechanism to track and adjust indicators would help ensure that

- core knowledge and skills are situated in the optimum grade with all prerequisites in place;
- no significant gaps in core content appear in strands across the grades;
- content evolves in cognitive complexity from one grade to the next;
- language is precise enough for teachers to understand the level of performance expected of students;
- anchors are organized as tightly as possible so redundancies are eliminated and reporting categories have a consistent grain size; and
- opportunities for integration across related areas of the language arts curriculum — writing, speaking and listening, media study, and research — are made more visible.

This type of matrix also could serve as a guide for the state’s next round of test development — a useful tool for formulating test specifications and developing sample tests.

As noted previously, Pennsylvania is currently developing a cross-grade matrix and will make it available to teachers as soon as possible.

✓ **Implement the state plan to provide examples by means of an Item Sampler and Item Bank of released test items.**

Because language used to describe the Assessment Anchors and Eligible Content statements is so similar across grades 3 through 11, the burden of showing a progression in cognitive demand falls on example items. In its review of the Pennsylvania’s summer 2004 edition of the reading document, Achieve urged the state to reconsider its “example items” in terms of their quality and their match to the statements of eligible content. In response, the state removed the example items from the Assessment Anchors and decided instead to provide example items by means of an Item Sampler and Item Bank of released assessment items on its Web page.

While the state’s intended Item Sampler and Item Bank should add clarity and rigor to the reading document, Achieve strongly recommended, in addition, that sample items be accompanied by at least one example per grade of the kind of text students are expected to comprehend to help target cognitive demand. Passage topic, length, content, organization and complexity are essential to estimating an assessment item’s worth and difficulty.

In terms of text complexity, the document *About the Reading Assessment Anchors* states that “the expectation is that the level of texts themselves will grow in complexity” and that the anchors “vary somewhat by listing such texts” (ii). Perhaps it is the case that the materials are still in development, but the materials include no listing of texts, except for the inclusion of the phrase “appropriate to grade level” added to the end of anchors. Genres of texts often are listed, but including stories, folk tales and poetry at grade 5 (R5.A.1) and short stories, excerpts from novels, legends, and poetry, including limerick and free verse, at grade 8 (R8.A.1) does little to indicate an increase in complexity of texts. Fifth graders are reading novels, although typically referred to as “chapter books,” and 8th graders may be reading legends. To clearly indicate a progression in expectation of complexity, the state needs to provide an additional way of explaining what “appropriate to grade level” entails. Pennsylvania could consider referencing their reading lists if they consider these lists to be indicative of the quality and complexity of the reading passages to be included on the assessments. The state also could choose to describe the dimensions of complexity in texts, as developed by Achieve and presented as an appendix to our technical report.

Achieve’s benchmark states have arrived at different strategies for communicating the level of text students are expected to comprehend. Massachusetts offers a series of sample grade-level reading passages in their standards, while New Standards provides a sample reading list. Indiana responds to the need for precision by including examples in the grade-level standards themselves:

Indiana Grade 5 Standard “Narrative Analysis of Grade-Level-Appropriate Text,” **Indicator 5.3.4:** *Understand that “theme” refers to the central idea or meaning of a selection and recognize themes, whether they are implied or stated directly.*

***Example:** Describe the themes in a fictional story, such as *A Wrinkle in Time* by Madeleine L'Engle, in which the themes of courage and perseverance are explored as the children in the story go on a dangerous mission in search of their scientist father.*

Adopting any one of these strategies will make the expected level of student performance far more concrete.

In response to Achieve's recommendations, Pennsylvania has decided to remove the list of genres included in the Assessment Anchors and is developing examples to illustrate the complexity of text students are to comprehend at each grade level.

- ✓ **Develop vertically aligned grade-level standards for grades K, 1, 2, 4, 6, 7, 9, 10 and 12 when the state next revises its academic standards in English language arts, and raise the level of rigor over time.**

Since 1999, when Pennsylvania last revised its standards, significant research has been conducted in language arts — at one end of the spectrum in the area of early literacy, and at the other end in the area of expectations of higher education and the workplace. It therefore makes sense for the state to revise its academic standards to reflect the latest research and to develop a full complement of standards for grades K–12. (Currently, state standards are in place only for grades 3, 5, 8 and 11.) Updating the standards also would provide an opportunity for Pennsylvania to increase the rigor of its expectations to match those of Achieve's benchmark states and our American Diploma Project. For example, the state may want to raise its demands by including allusions and analogies in its vocabulary section; adding the assessment of a literary text in terms of its historical, social, and/or political context; and asking students to explain how a writer of non-fiction purposefully chooses language in developing meaning and establishing tone and style. In the end, Pennsylvania wants to have a ladder of expectations that takes students from the early years and makes them college and work ready by graduation.

MAJOR FINDINGS: ALIGNMENT OF ASSESSMENTS TO ASSESSMENT ANCHORS AND ELIGIBLE CONTENT IN READING

Achieve carried out a detailed review of the alignment between Pennsylvania's reading assessments at grades 3, 5, 8 and 11 and the corresponding Assessment Anchors and Eligible Content statements as contained in the September 2004 edition (and not the November 2004 edition posted on the Pennsylvania Web site).

Each of the grade-level assessments that were reviewed included five to six reading passages, followed by a set of multiple-choice and constructed-response reading items related to that text.

The form and format of the assessments that were reviewed are shown in the table below:

FORM AND YEAR	NUMBER AND FORMAT OF ITEMS	NUMBER OF POINTS	NOTES
CTB 1 <i>Gr. 3</i> / 9-13-04 Received 9-15-04	40 multiple choice 2 brief constructed response	40 6	Five reading passages
DRC <i>Gr. 5 Core</i> 2005 Received 9-02-04	40 multiple choice 4 brief constructed response	40 12	Five reading passages
DRC <i>Gr. 8 Core</i> 2005 Received 9-02-04	44 multiple choice 4 brief constructed response	40 12	Six reading passages
DRC <i>Gr. 11 Core</i> 2005 Received 9-15-04	44 multiple choice 4 brief constructed response	40 12	Five reading passages

STRENGTHS OF THE ASSESSMENTS

- **Pennsylvania's tests at grades 3, 5, 8 and 11 are strongly aligned to the state's new Assessment Anchors and Eligible Content statements.**

Achieve remapped some items to Eligible Content statements where we judged that a better match existed than the one the test developer originally designated. The result is that the content and performance alignment between Pennsylvania's Assessment Anchors and Eligible Content statements and the grades 3, 5, 8 and 11 assessment items is strong. As shown in the table on the next page, at each grade a majority of items received scores of "2" or "1b" for content centrality and for performance centrality. These scores indicate items are either fully or partially aligned — due to the compound nature of the Eligible Content statements — to the content of those statements.

GRADE LEVEL	MAPPED ITEMS Number/Percent	CONTENT CENTRALITY SCORES Number/Percent				PERFORMANCE CENTRALITY SCORES Number/Percent			
		2	1a	1b	0	2	1a	1b	0
3	42	31	0	10	1	41	0	0	1
	100%	74%	0	24%	2%	98%	0	0	2%
5	44	28	0	10	6	21	0	12	11
	100%	64%	0	23%	13%	48%	0	27%	25%
8	44	32	0	11	1	24	4	11	5
	100%	73%	0	25%	2%	55%	9%	25%	11%
11	44	34	0	10	0	29	0	12	3
	100%	77%	0	23%	0	66%	0	27%	7%
TOTAL	100%								

Scores of “1a” for content or performance centrality reflect standard statements that are too general or vague to show clear alignment with assessment items. The above data show that many of the revisions that Pennsylvania made to its Assessment Anchors since Achieve’s preliminary review in fall 2003 have clarified the anchors and made the statements of eligible content more focused and specific. The Pennsylvania reading assessments at grades 3, 5, 8 and 11 received no item scores of “1a” for content centrality, and only four items out of 174 total items across the four grade levels received item scores of “1a” for performance centrality.

- **Pennsylvania’s reading passages increase in cognitive demand across the grade levels and have a good balance of genres.**

Reviewers found that the reading passages at grade 5 were more challenging than those at grade 3; those at grade 8 were more challenging than those at grade 5; and those at grade 11 were more challenging than those at grade 8.

Achieve reviewed 21 reading passages across the four grade levels. The selections covered a range of genres, including poetry, short story, interviews and other informational texts, and this variety of genres allowed the state to address many of its Assessment Anchors.

AREAS FOR IMPROVEMENT

- **Many statements of Eligible Content are not assessed in Pennsylvania’s current reading tests.**

Of the 122 total statements of Eligible Content across all four grade levels, 54 statements of Eligible Content were assessed by the grade-level reading assessments and 68 statements were not assessed. The table on the next page shows the number of statements assessed and not assessed at each grade level.

**Number of Eligible Content Statements Assessed and Not Assessed:
Grades 3, 5, 8 and 11**

	GRADE 3	GRADE 5	GRADE 8	GRADE 11
Number/Percent Assessed	16/55%	13/42%	12/39%	13/42%
Number/Percent Not Assessed	13/45%	18/58%	19/61%	18/58%
TOTAL	29/100%	31/100%	31/100%	31/100%

While every test is a sampling of a larger domain of content, Achieve recommends that the state include items that map to a higher percentage of the Eligible Content statements, particularly because these statements are based on a set of Assessment Anchors — a subset of Pennsylvania’s comprehensive academic content standards.

Although the reporting categories for student and school scores are based on the larger Assessment Anchors and not the Eligible Content statements, these statements nevertheless are written at an appropriate grain size for test items. Teachers are quick to digest what content is actually assessed and, in a high stakes environment, tend to modify their instructional program to help students score well. It is therefore important from an instructional stance that as many Eligible Content statements be assessed as is feasible, given testing time constraints and reporting requirements. The following observations may be helpful to the state in improving coverage.

Some Eligible Content statements are assessed at only one grade.

Some of the statements of Eligible Content that were not assessed were not assessed at grades 5, 8 or 11. For example, there were no items at grades 5, 8 or 11 that assessed student knowledge of multiple-meaning words; synonyms; affixes and roots; graphs and charts; headings; or bias, stereotypes, and propaganda. The state may want to consider what the failure to represent these statements on the assessment signifies.

Some Eligible Content statements may be overassessed.

Of the 174 total items across all four grade levels, 58 items (33 percent) were mapped to the two statements of Eligible Content that dealt with main idea and details in fiction and non-fiction texts. Twenty-nine items (17 percent) were mapped to the Eligible Content statement that related to literary elements such as plot, character, setting and theme. The state will want to consider whether this range of items appropriately reflects its Assessment Anchors. In addition, the state may want to consider the nature of these statements of Eligible Content and whether their content is broader than other narrower statements. The state might want to break down each of the statements to which so many items were mapped into smaller pieces — main idea, for example, could be separated from details.

Some statements of Eligible Content were not assessed because appropriate passages were not included.

The state should re-examine the nature of the passages that deal with informational text. Some Assessment Anchors could not be assessed because the texts included on the assessments were not persuasive or argumentative in nature and did not include graphic elements such as maps, charts or graphs.

Some statements of Eligible Content require further clarification.

Some statements of Eligible Content, particularly those that were not represented by items on the assessments, were unclear to reviewers and may not have been assessed for that reason. For example, what does the following mean: *Identify examples of text that support its narrative, informational, persuasive, or instructional purpose* (R8.A.2.6.2)? Other statements may not have been addressed, as it was unclear how they could be included in a large-scale, on-demand test, such as R5.B.2.1.2: *Identify lines from a poem where a definite meter is discernible*.

Following through on Achieve’s advice, Pennsylvania has made the following decisions with its test contractor. It has developed a plan to ensure that every anchor is tested at least every four years — understanding that some anchors need to be assessed every year, such as “main idea.” Pennsylvania also has developed guidelines for the kind of reading passages that must be included so the state has the potential to assess key anchors, such as those dealing with comprehending informational text.

- **Pennsylvania should ratchet up the cognitive demand of its reading anchors, content statements and tests — particularly at grade 11 — over time and in concert.**

We live in an information age and are part of a global economy. This reality means it is no longer the case that students bound for college need more academic training than those bound for work. In fact, Achieve’s American Diploma Project found that college professors and employers were in agreement that high school graduates must exit with advanced critical reading and analysis skills to be college and work ready. These skills are built grade by grade, and a state’s standards and assessments must steadily spiral in demand if students are to meet with success. For this reason, Achieve urged Pennsylvania to refine its reading Assessment Anchors and Eligible Content, bolstering these with examples of text and recommended reading lists that make the trajectory clear.

The state also will want to review the overall rigor of its test items. In judging the level of cognitive demand of reading assessments, Achieve evaluates passages and related items. To capture the kind of mental processing an item requires, we assign a Level of Demand score from 1 to 4 as described below.

<u>Level 1</u>	Items that require little beyond simple recall or identification.
<u>Level 2</u>	Items that demand some level of inference going beyond recalling or reproducing a response.
<u>Level 3</u>	Items that require the student to make an interpretation (often constructed-response items).
<u>Level 4</u>	Items that describe an extended, constructed-response task that requires investigation and synthesis. (These items are not generally found on large-scale, on-demand tests.)

The table on the next page shows the levels of demand for items on the grades 3, 5, 8 and 11 reading assessments (based on Achieve-recommended mapping). The percentages in the table reflect the total number of those items able to be rated for level of demand.

Progression of Scores of Level of Demand across Grades 3, 5, 8 and 11

READING ASSESSMENTS		LEVEL OF DEMAND (Number of Items & Percentage per Level)				
Grade	# of Items	1	2	3	4	nr*
3	42	9/21%	29/69%	2/5%	0/0%	2/5%
5	44	11/25%	23/52%	4/9%	0/0%	6/14%
8	44	15/34%	21/48%	4/9%	0/0%	4/9%
11	44	9/20%	27/61%	4/9%	0/0%	4/9%

* Items rated “0” for content centrality, performance centrality or source of challenge are not rated (nr) for level of demand.

The proportion of Level 1 items on the grade 8 assessment is a cause for concern, given that students must be prepared to handle demanding informational text in high school science and social studies courses, as well as more complex literature in their English courses. Pennsylvania may want to increase the percentage of Level 3 items over time.

The grade 11 assessment includes fewer Level 1 items and more Level 2 items, and such a pattern would be expected with a higher-level assessment. The four items that received Level 3 scores are the constructed-response items on the grade 11 assessment. The state may want to consider increasing the number of constructed-response items on the grade 11 assessment to increase the level of demand. Other ways to increase the level of challenge for grade 11 students would be to increase the difficulty of the passages and increase the level of performance expected on the multiple-choice items. Pennsylvania could look to the ACT exam as an example of a high school assessment where reading passages and items may be more challenging than those that appear on Pennsylvania’s grade 11 assessment. An example of an ACT passage and related items is presented as an appendix to the technical report that Achieve prepared for Pennsylvania.

RECOMMENDATIONS FOR IMPROVEMENT

- ✓ **Determine a goal or target in terms of the percentage of statements of Eligible Content that should be assessed on its tests.**

As noted above, currently at grades 5, 8 and 11, the majority of statements are not assessed. The coverage seems inadequate given that the Eligible Content statements are written as testable content for the PSSA (unlike the academic standards, which include content that would not be tested on a statewide, on-demand assessment).

- ✓ **Adjust the balance of items among the different Assessment Anchors to ensure that the content the state deems most important is being sufficiently — but not overly — assessed.**

A second look at the balance of items mapped to different statements is in order. The state may affirm that the number of items currently assessing “main idea and details” is on target, as these are the most important reading skills for students. On the other hand, a significant number of Eligible Content statements were not assessed, and it may be that the state concluded these areas are not important enough to be assessed. If that is the case, they may not be worth including in the Assessment Anchors. Alternatively, the state may

realize that those statements that are not assessed for reasons of being more challenging *should* be assessed.

- ✓ **Revise the wording of identified Assessment Anchors and Eligible Content statements to improve their clarity.**

It is crucial that educators know what kinds of knowledge and performance are expected by reading the Assessment Anchors. In its technical report, Achieve noted those anchors and content statements that require revision.

- ✓ **Specify to contractor(s) the types of reading passages needed to construct items that clearly assess specific anchors and related content.**

If the state feels that it is important to assess students' ability to comprehend persuasive and argumentative passages and to analyze and evaluate tables, charts and graphs, then passages that lend themselves to asking these types of items are needed.

- ✓ **Increase the level of demand of the assessments (items and passages) across the grade levels over time.**

The Assessment Anchors do not change dramatically across the grade levels. To increase the challenge for grade 8 and grade 11 students, the state may want to consider increasing the reading level of the texts and/or increasing the level of demand of the items (for example, by including additional open-ended items). Currently, the texts on the assessments show a progression in their level of difficulty, but the passages may represent the lower levels of reading appropriate to each grade, and the percentage of items that assess advanced thinking skills is sometimes low.

For the earlier grades, it is important to be responsive to issues of universal design and fairness in testing and provide texts that are at grade level (as determined by experts), yet still reasonably robust. Room for the most advanced students to demonstrate their performance on these elementary-level assessments can still be provided by including a sufficient number of items that tap students' ability to think critically. However, at grade 8, and again at grade 11, it would seem appropriate to evaluate whether students are capable of comprehending the full range of texts they will encounter in high school and then in postsecondary education and the workplace.

MAJOR FINDINGS: REVIEW OF PENNSYLVANIA’S ASSESSMENT ANCHORS AND ELIGIBLE CONTENT IN MATHEMATICS

INTRODUCTION

In our review of Pennsylvania’s preliminary 2003 document in mathematics, Achieve recommended the state give special attention to four issues: focus, redundancy, coherence and balance. In making revisions, Achieve advised Pennsylvania to employ a strategy of pruning and grafting to keep the amount of content manageable. Achieve made the following specific suggestions:

- **FOCUS:** Decide on an emphasis for each grade level and then prune that grade level’s standards accordingly. For example, place more emphasis on number concepts, measurement and geometry in the earlier grades (3–5); on geometry, algebra and statistics in the middle grades (6–8); and algebra, statistics, probability, calculus, and geometry and trigonometry in the later grades (grade 11). By extension, content that is more applicable to a different grade level should be grafted there. In carrying out this task, it will be important to group concepts and skills that should be taught together to increase student comprehension and retention of significant content.
- **REDUNDANCY:** Find the concepts, ideas and standards that are repeated inappropriately across the grade levels and prune these.
- **COHERENCE:** Improve the way the content statements are seen as elaborations of the Assessment Anchors. Add examples to signal the level of rigor expected and make the language more precise.
- **BALANCE:** Improve the balance of content statements devoted to procedural understanding, conceptual understanding and problem-solving expectations.

Achieve found that overall Pennsylvania’s revisions to its 2003 draft in mathematics succeeded in sharpening the focus on essential mathematics at each grade level, reducing redundancy, strengthening coherence (by better connecting statements of Eligible Content to the related anchor and improving their specificity) and making some adjustments to balance (the attention given to conceptual understanding, procedural knowledge and application). However, Achieve also concluded that more substantive work was required for the standards document to reach a level where the state could be assured its students are programmatically on track to attain proficiency at each grade level.

Unlike reading, to which the state was able to make immediate adjustments, Pennsylvania concluded Achieve’s recommendations (summer 2004) for improving the Assessment Anchors and Eligible Content statements in mathematics would involve major adjustments in content that the state could not implement in time to affect its 2005 or 2006 tests. Instead, the state plans to implement most of Achieve’s recommendations in time to affect its 2007 tests in mathematics. This means, in practice, that the version of Pennsylvania’s Assessment Anchors and Eligible Content statements on the state’s Web site (November 1, 2004) will remain in place until after the close of the 2005 test administration period.

Achieve was not able to undertake a second comprehensive review of the state’s final revisions to the Assessment Anchors and related Eligible Content statements. We did, however, conduct an abbreviated review that indicated the modifications with which we were most concerned are in place for the 2007 assessment and should result in better instruction and assessment across the state.

STRENGTHS OF THE ASSESSMENT ANCHORS

- **For the most part, the Eligible Content statements are written in clear and measurable language and, on average, have an equivalent grain size or level of specificity.**

The Eligible Content statements typically — and rightly — include tasks that describe the result rather than the process of learning. Verbs that focus on the result of learning (e.g., identify, calculate, locate, explain, solve, analyze, etc.) readily translate into testable items. Most of the Eligible Content statements also are detailed enough to address specific topics or concepts.

Although the anchor statements are too broad to be useful for developing assessment items, the Eligible Content statements usually succeed in supplying the specificity required to develop assessment items, as the following grade 5 example illustrates:

ASSESSMENT ANCHOR	
M5.A.1 Demonstrate an understanding of numbers, ways of representing numbers, relationships among numbers and number systems.	
	ELIGIBLE CONTENT
M5.A.1.4 Use simple applications of negative numbers (number line, counting, temperature). <i>Reference: 2.1.5.F</i>	M5.A.1.4.1 Identify negative numbers on a number line (greater than or equal to –20). M5.A.1.4.2 Identify negative numbers on a thermometer (°C or °F).

Pennsylvania Department of Education
Assessment Anchors and Eligible Content

Math Grade 5 – Page 37
Web site 11/1/04

Another effect of the detail is to make the grain size of the Eligible Content statements more comparable across the grades.

Achieve noted that some Eligible Content statements remained insufficiently precise in Pennsylvania’s summer 2004 document. For example, in revising Eligible Content statements at grade 11, the state should label those that are “non-calculator” (meant to be responded to without the aid of a calculator).

In addition, the state should revisit the phrase “in various contexts.” Teachers typically interpret this to mean problems will be framed in a real-world situation rather than just involve symbol manipulation. Most of the cited examples Achieve originally reviewed were not situational — a potential source of confusion.

Achieve suggested ways to clarify other statements and included these as line edits in an appendix to our technical report, which was sent to the state in November 2004.

-
- **By specifying the grade-level knowledge and skills that are subject to state assessment, the Assessment Anchors and Eligible Content statements will assist teachers in lesson planning and classroom assessment.**

Knowledgeable teachers can readily figure out what lessons they need to provide their students to prepare them for assessments based on the Eligible Content statements. These statements also should help teachers evaluate the usefulness of their textbooks in supporting lessons. While the statements in the summer 2004 edition were specific enough to guide lesson development, Achieve recommended the state focus topics more sharply to encourage the design of rich mathematical units, rather than disconnected lessons on many small, isolated topics. For example, teaching ratio, proportion, percent, rates, scaling and similarity as an integrated unit over several months would connect these mathematical ideas for maximal effect. Achieve was pleased to see that Pennsylvania has clustered these topics at grade 7 in the final 2007 version of the mathematics documents.

Achieve also observed that understanding rates of change with graphs and tables leads to understanding slope; linearity, including the equation $y = mx + b$; and proportionality, $y = mx$. Similarly, the Pythagorean Theorem relates to roots and irrational numbers as well as to distance on a coordinate grid. Pennsylvania rightly includes the Pythagorean Theorem for study at grade 8 but revisits it at grade 11, along with irrational numbers.

This “big idea” approach to instruction, as opposed to a march through a multitude of smaller topics, is more beneficial to student learning and retention.

- **The Assessment Anchors are easy to follow and cross-reference.**

The introductory overview, “About the Mathematics Assessment Anchors,” is brief and clear, answering most of the questions one might have about what the anchors were intended to do and how they are organized. In reorganizing the content in the state’s academic standards by regrouping 11 standards into five major strands, Pennsylvania improved the focus, organizational structure and usefulness of the mathematics document, particularly for elementary teachers who are primarily generalists. This reorganization will support the state’s goal of centering on the durable knowledge students must internalize if they are to end up being mathematically literate.

AREAS FOR IMPROVEMENT

- **Achieve’s review of the state’s mathematics document in summer 2004 indicated that the amount of content to be taught and learned effectively was not manageable — particularly at grades 8 and 11.**

Achieve cautioned that the summer 2004 document was overstuffed, packing in too much content to be taught in a balanced way to ensure students’ full understanding. If students are to develop genuine understanding (i.e., acquire the related procedural, conceptual and strategic knowledge underlying the mathematics) — and not just depend on memorization of facts and the mechanical application of algorithms — then time must be provided for them to do so.

To communicate Achieve’s concern that an overload of content can undermine rigor and depth of understanding, we offer the following illustration: Imagine a unit of study

for grade 4 that develops the concepts of “perimeter” and “area” for both regular and irregular figures such as footprints or leaves using color tiles and grid paper. The unit calls for students to develop an investigation of what happens when varying one parameter (perimeter or area) while holding the other constant. The unit includes strategies for helping students develop the appropriate formulas, tables and graphs to show the relationship of perimeter and area related to a side. Importantly, the unit provides instruction not only on the skills involved but also on the inherent conceptual understandings. Furthermore, it includes problem solving as an investigative technique for exploration and deepening understanding of the concepts. A rich unit such as this could take three weeks — time that is not available if there are 49 grade 4 topics to cover before the state test. (And, if the state limits its expectations to regular figures, such as the perimeter and area of squares and rectangles, such a unit likely would not be taught at all.)

Achieve made specific suggestions for grade-level changes in its technical report that would result in a more manageable distribution of topics in grades 3–8. (Grade 11 is a somewhat different situation and will be addressed separately.) For example, we noted that one way to trim content and gain a sharpened focus is to look for single content expectations that are not connected to any other concepts at a given grade level and move these to a grade where they can be connected to other learning. For example, likelihood (grades 3 and 4) and combinations (grade 4) are two such topics that would be better placed in a later grade, and ratios (at grade 6) is another topic that appeared to be somewhat isolated.

The state acted on these recommendations. Its final version indicates likelihood and combinations have been moved to grade 5. Ratio has been moved to grade 7, where it can be taught in an integrated unit of study. Pennsylvania also took Achieve’s advice and pruned 59 Eligible Content statements in grade 8 down to 43 in its final 2007 version.

- **The progression of the Eligible Content statements in mathematics needs to be fine tuned.**

The state made some progress in eliminating redundancies in its summer 2004 document, and taken as a whole, the Eligible Content statements describe an adequate progression of more complex understanding across the grades. However, more can be done to streamline the content at each grade level to underscore emphases, avoid unnecessary inconsistencies and provide enhanced opportunities for students to learn the key mathematics at their grade level. Achieve identified examples of Assessment Anchors, Eligible Content statements and illustrative examples that would be better placed at a different grade level because of inconsistency with the related standard. For example, line graphs are inappropriately included at grade 4 and better placed at grade 8. We also noted instances in which there was a mismatch among related ideas and expectations. For example, if 4th graders are learning about eighths as fractions (as is specified in Number), they should be measuring to the nearest eighth of an inch (instead of the nearest one-fourth inch, as is currently specified in Measurement). Finally, sometimes the state’s example items were mismatched to grades in terms of the level of challenge they contained.

Achieve recommends the state develop a content matrix to verify progression and improve the utility of its document. A matrix is a continuum chart that organizes topics across grades in rows of similar content to delineate the progression of knowledge and

skills. Such a continuum will enable districts to chart the development of each topic and teachers to see what their students were expected to learn before and what they will be expected to learn after they leave their class.

A related improvement needed to make such a continuum possible is making the numbering system more consistent. For example:

- A1.2 is “Compare quantities and/or magnitudes of numbers” at all grades except at grade 5, where it is listed as A1.3.
 - Number Theory is A1.3 at grade 4, A1.6 at grade 5, A1.4 at grade 6 and A1.2 at grade 11.
 - M8.A1.1.2 should correlate with M6.A1.1.3 and M11.A1.1.3 — all of which deal with exponential form and scientific notation.
 - M8.A1.1.3 and M11.A1.1.2 both cover square roots.
- **Pennsylvania omits some essential ideas or fails to place them in optimal grades, as required to build students’ understanding across the grades.**

Achieve identified a few essential ideas that were not contained in the state’s summer 2004 mathematics document. For example, the commutative, associative, distributive, identity and inverse properties, typically addressed at grade 6, were not included. However, the state corrected the omission, with the exception of the inverse property, in its 2007 version. Common irrational numbers and proportionality — as a special case of linearity — should be included in grade 8 but also were omitted.

The state should reconsider the placement of some other core concepts. The number line is introduced in grade 4, whereas it should be introduced at least by grade 3. Equivalency of fractions is included as part of the content at grade 6, but it is far more appropriately placed in grade 4. While direct and inverse variation are included at grade 11, they should be studied no later than grade 8, where they logically fit and where they support study in physical science.

- **Pennsylvania’s strand in geometry as expressed in its Assessment Anchors and Eligible Content statements is relatively weak.**

As noted earlier, Achieve recommended that geometry be given more emphasis in every grade. One major function of geometry in the K–8 curriculum is to provide a concrete and familiar setting in which children can learn to do mathematics — to work with definitions, conjectures and proofs. The study of geometry is important not only for its practical applications in many occupations but also because it trains students to formulate and test hypotheses and to justify arguments in formal and informal ways. Employers stress the importance of understanding geometry, especially in the construction and manufacturing industries. Achieve’s reviewers were concerned that this important area of study in mathematics was not well developed across the grades in the Assessment Anchors and Eligible Content statements.

Achieve’s abbreviated review of the state’s final document indicates Pennsylvania has taken steps that have somewhat strengthened this important strand.

- **The content of the grade 11 anchors requires augmentation, especially in the area of Algebra II.**

Achieve’s American Diploma Project study revealed that high school graduates need a command of mathematics through and beyond Algebra II, to include study of data analysis and statistics, if they are to be on track to earn postsecondary degrees and be successful in today’s workplace. Sobering data back up that assertion. Our study revealed that 75 percent of high school graduates enroll in postsecondary education programs. Of those, nearly 30 percent are placed immediately into a remedial college course. Even more alarming, most students who take remedial courses fail to earn either an associate’s or bachelor’s degree. Today’s work requirements are similar to postsecondary expectations. Mathematical knowledge and proficiency are a given for employment in fields such as computer science, operations research and management science. It therefore makes good sense for Pennsylvania to build content beyond Algebra I and geometry into its Assessment Anchors and Eligible Content statements at grade 11. Recent changes to the SAT indicate it also is moving in the direction of including Algebra II content, as the ACT already does.

Specifically, Achieve recommends the state enhance its treatment of quadratics. We did not find mention of transformations of quadratics (e.g., students should be able to describe the graphic change that occurs when $y = 2x$ becomes $y = 2x + 8$).

We also suggest the state add to or augment the following Eligible Content statements that are related to the study of quadratics.

2003 ELIGIBLE CONTENT STATEMENT	REVIEWER’S RECOMMENDATIONS	2005/2007
M.11.D.2.1.5: <i>Solve quadratic equations using factoring.</i> (integers only — not including completing the square or the quadratic formula).	<i>Change</i> to “...using factoring or the quadratic formula.”	No change
M.11.D.4.1.2: <i>Graph linear functions in two variables.</i>	<i>Remove</i> the word “linear” and <i>replace</i> with “quadratic/exponential.”	No change
M.11.D.4.1.4: (does not exist in 2003 Eligible Content statements)	<i>Add</i> “Solve a system involving a quadratic or exponential function with a linear, quadratic or exponential function.”	Not added
M11.D.4.1.3: <i>Determine the maximum or minimum of a quadratic function.</i>	<i>Retain.</i>	Deleted in 2007

In addition, Achieve recommends that the state augment its anchors and Eligible Content statements in the area of statistics and incorporate the study of exponential growth and decay.

-
- **Pennsylvania should develop a full complement of example items to clarify statements of Eligible Content and make the level of cognitive demand expected as precise as possible.**

In its summer 2004 review, Achieve noted that the example items the state had included in its mathematics document to elucidate the Eligible Content statements were uneven in quality and utility. These deserve attention because they are typically the first place readers look for clarification of what an anchor or Eligible Content statement really means.

Achieve understands the state intends to eliminate “example items” in the body of the mathematics document and instead provide sample items and a bank of released items from previous tests on its Web site. This will provide an excellent opportunity for the state to deepen teacher knowledge with well-considered choices of items.

Some of the state’s examples were outstanding. For example, a number of the multiple-choice items had thought-provoking distracters, in that some or all of the incorrect answer choices were based on specific (often common) misconceptions. Teachers are not generally trained to look carefully at the wrong answers in a multiple-choice question. Yet if they did, every well-crafted multiple-choice question would give them useful information about the ways in which their students were likely to misunderstand the mathematical concepts they were trying to learn. In light of this characteristic of multiple-choice items, Achieve recommends that Pennsylvania include brief comments about the distracters in its example items, highlighting the misconception or manipulation error each distracter represents. This addition would enhance readers’ understanding of both the items and the mathematical concepts they assess and would be extremely helpful to teachers.

Some other choices of examples were less successful — in some instances, examples were matched to Eligible Content statements that they did not measure well. Another concern was that some Eligible Content statements were illustrated by sets of two or three items while others were not accompanied by any.

Because Pennsylvania intends to include open-ended items as part of the state assessment, it should include examples of open-ended tasks for each content strand — Numbers and Operations, Measurement, Geometry, Algebraic Concepts, and Data Analysis and Probability — in the Assessment Anchors. These examples should help teachers understand the problem-solving strategies and explanations and the applications of conceptual understandings students are expected to use or produce. The current version of the Assessment Anchors may lead teachers to believe that the multiple-choice questions on the PSSA are paramount, and in response, they may end up teaching a curriculum of small facts and procedures.

Developing sample items and a bank of released items has the potential to provide substantive help to teachers and students. We strongly encourage the state to include explanatory notes as mentioned above and a full complement of open-ended tasks to gain maximum effect from its efforts.

-
- **While some attention has been given to the issues of balance of content, the current Assessment Anchors and Eligible Content statements overemphasize procedural knowledge at the expense of conceptual knowledge, reasoning and problem solving.**

Mathematical proficiency depends on many factors other than balanced knowledge of content. These include procedural fluency, conceptual understanding and flexible reasoning. Achieve urges Pennsylvania to aggressively develop conceptual understanding and mathematical reasoning from the earliest grades to ensure progression of mathematical understanding.

The examples that currently accompany each Assessment Anchor and its corresponding Eligible Content statements help illustrate what that anchor means, but they also convey an unbalanced emphasis (as they did in the 2003 draft) on procedural knowledge. Pennsylvania's reasonable effort to define and limit what can be asked on a test is a worthy undertaking. A major shortcoming, however, is that nearly everything that is usefully specific is couched in procedural rather than conceptual terms.

The majority of the statements in the current draft of Assessment Anchors and Eligible Content statements require that students perform a routine, well-practiced procedure. In the fall 2003 review, Achieve estimated that the percent of content statements stressing procedural — as opposed to conceptual or problem-solving aspects of mathematics — ranged from approximately 86 percent at grade 8 to 97 percent at grade 6. Even though those percentages are slightly lower in the summer 2004 draft, they are still unbalanced. Mathematically strong teachers can easily fill in these omissions to create a rigorous program. However, many K–8 teachers who are unsure of their own mathematical proficiency may read these standards literally and teach them narrowly, leaving students with only procedural skills and little conceptual understanding.

In addition, the state's summer 2004 draft does not address explicitly the problem-solving aspect of the expectations. The only mention of the Problem Solving (2.5) and Reasoning (2.4) standards is in the introduction, which states that these standards are not specified in the Assessment Anchors because the anchors deal with content, not process. Pennsylvania clearly recognizes the importance of reasoning and problem solving and reports results for open-ended items that emphasize problem solving separately. Achieve recommends that the state underscore the centrality of these skills and make their connection to the anchors more explicit. Good examples that rise above the level of procedure could overcome this deficiency.

- **The level of rigor of the Assessment Anchors and Eligible Content statements is somewhat below the rigor of Achieve's benchmarks.**

Pennsylvania's treatment of topics sometimes appears to be less rigorous than that found in the benchmark documents. However, Achieve has found that it often is very difficult to determine the level of rigor absent illustrative tasks or sample problems. When revising its academic standards, the state will want to take advantage of Achieve's emerging K–8 Benchmark Standards in Mathematics. These benchmarks are meant to guide states' K–8 standards so they will be of sufficient rigor to prepare students for a demanding, four-year high school sequence. By the same token, Achieve's American Diploma Project's end-of-high-school benchmark standards will be helpful in developing course-level standards for high school.

RECOMMENDATIONS FOR IMPROVEMENT

Accelerate the state plan to make sample items and an item bank of released items available on Pennsylvania's Web site.

There are a number of compelling reasons for the state to accelerate its plan:

1. Illustrations or sample problems in mathematics are necessary to convey the intended cognitive level of Assessment Anchors and Eligible Content statements to teachers as they prepare their instructional program. Sample items telegraph the real level of proficiency students are expected to demonstrate. An advantage of the Web is the ease in which the state can add or reposition items as it raises the rigor of its expectations.
 2. Schools are under enormous pressure not only to raise the level of mathematics proficiency for all their students but also to close achievement gaps between subgroups of students who have been historically underserved by the education system. Teacher knowledge is key to student success. Teachers who are new to the profession or underprepared in their understanding of mathematics and/or mathematical pedagogy are best supported with explicit standards and examples. Pennsylvania has taken a significant step in this direction by culling a subset of knowledge and skills to be the focus of instruction and testing. Elucidating these with annotated sample items will help teachers figure out exactly what they need to know and the repertoire of approaches they need to have to help struggling students succeed.
 3. Students should see mathematics as a holistic, coherent discipline. Again, this is optimally done through supplying concrete examples that show significant relationships among major strands — for example, Geometry and Algebra — and demonstrate how the knowledge of two seemingly disparate concepts can be brought to bear on a problem.
- ✓ **Make public the revised Assessment Anchors and Eligible Content statements on which the 2007 assessments will be based as soon as possible.**

The new Assessment Anchors and Eligible Content statements will go a long way in improving instructional practices in mathematics across grades 3–11. Once the administration window closes on the 2005 tests, Pennsylvania should immediately post the final edition of the Assessment Anchors and Eligible Content statements. Two issues drive the urgency for immediate action. The first is curricular: School districts will be anxious to adjust their expectations and instructional materials at each grade level to reflect the increased focus of the revisions and the more targeted scope and sequence. However, districts will have to make a two-step adjustment, first by adjusting their curriculum to conform to the state's 2005 Assessment Anchors and Eligible Content statements, especially in the grades not currently tested (4, 6 and 7), and then by undertaking a more substantial realignment to respond to the state's fundamental regrouping of its essential content effective for the PSSA in 2007. The second issue is financial: Some school districts are likely poised to purchase instructional materials in spring 2005 — a huge financial investment that typically has a five-to-seven-year lifespan. Knowing what the 2007 PSSA demands will help school districts make judicious selections.

Achieve therefore recommends the state take a long view and adopt a proactive stance by publishing the 2007 expectations immediately after the 2005 tests have been administered. This publication should include footnotes about the discrepancies between the 2006 and 2007 versions.

- ✓ **Construct a cross-grade matrix or continuum so that the progression of essential knowledge and skills can be readily tracked from grade to grade.**

To show progression of knowledge and skills across grade levels, the state should construct a continuum that uses overlapping spans such as 3–5, 5–8 and 8–11 once the Assessment Anchors and Eligible Content statements are released for the 2007 testing cycle. Arguments for constructing such a tool are contained in our recommendations under the Reading section of this report.

- ✓ **Develop a pacing chart to focus attention on the full complement of mathematics expectations in each grade level.**

Ideally, state tests should be administered at the close of the school year. Pennsylvania's testing window is currently a function of the date on which the Easter holiday falls, which triggers the start of the traditional spring break. In practice, this means the state's testing window fluctuates based on a floating holiday that can occur in a period from early March to late April. This means that the amount of instructional time teachers have to devote to content prior to the PSSA varies significantly from year to year. Even if the state had a fixed window for spring testing, there would remain a tendency for schools to cram before and let up after testing, leaving at least two months that may not be used to optimum advantage. To combat this proclivity, Pennsylvania may want to develop a "pacing chart" to highlight the mathematics that should be taught and learned in the weeks that follow the administration of state tests. A pacing chart would communicate that these weeks are critical in preparing students for the demands of the next grade's work.

- ✓ **Develop vertically aligned grade-level standards (or, in the case of high school, course standards) for grades K, 1, 2, 4, 6, 7, 9, 10 and 12 when the state next revises its academic standards in mathematics.**

Pennsylvania has not revised its academic standards since 1999. Experts in the field of mathematics have since paid considerable attention to issues of grade-level placement, the emphasis given to core content, and the kind of knowledge and skills students should acquire by graduation. When revising the academic standards, the state should take these latest developments into account.

The state's reorganized content strands are the right architecture for restructuring the academic standards in mathematics and for completing a full complement of K–12 grade-level standards or course standards at high school.

At one end of the educational continuum, foundation skills in grades K–2 are important to help ensure students' success in grade 3 mathematics (as well as in science). At the other end, it is quite clear that students in the United States require a greater command of mathematics to function successfully in the world in which they will live and work. Yet current data suggest students in the United States are falling farther behind students in high-performing countries. Achieve has undertaken two initiatives to help states improve

mathematics education. We are in the process of finalizing K–8 standards in mathematics, and under the auspices of our American Diploma Project, we have established benchmarks for graduation. Achieve urges Pennsylvania to make use of our emerging K–8 standards and our American Diploma Project benchmarks to update its K–12 standards.

MAJOR FINDINGS: ALIGNMENT OF ASSESSMENTS TO ASSESSMENT ANCHORS AND ELIGIBLE CONTENT IN MATHEMATICS

Achieve carried out a detailed review of the alignment between Pennsylvania's assessments at grades 3, 5, 8 and 11 and its Assessment Anchors and Eligible Content statements in mathematics.

STRUCTURE OF THE ASSESSMENTS

The PSSA 2005 provides individual student scores based on a common set of items and program scores based on a combination of the common items and matrixed items. Each student booklet also contains eight or nine field test items that were not evaluated by Achieve. Students take the mathematics tests in three sittings. Achieve conducted its Alignment-to-Standards (ATS) Review in September 2004, using an updated version of the Assessment Anchors for grades 5, 8 and 11 and a version dated 4/30/04 for grade 3.

FORM AND YEAR	NUMBER AND FORMAT OF ITEMS	NUMBER OF POINTS	RESOURCES PROVIDED TO STUDENTS*
CTB 1 <i>Grade 3</i> 9-13-04	54 multiple choice 2 open ended	54 8	Punch-out ruler, 7 inches on one edge, marked in fourths; 18.5 centimeters on other edge, marked in millimeters.
DRC <i>Grade 5</i> Core 2005	54 multiple choice 3 open ended	54 12	Punch-out ruler, 6 inches on one edge, marked in sixteenths; 15 centimeters on other edge, marked in millimeters.
DRC <i>Grade 8</i> Core 2005	54 multiple choice 3 open ended	54 12	Formula sheet with equivalents for common and customary units.
DRC <i>Grade 11</i> Core 2005	54 multiple choice 3 open ended	54 12	Formula sheet.

* Each assessment in mathematics contains items that are designated as "non-calculator"; calculators may be used with the remaining items.

STRENGTHS OF THE ASSESSMENTS

- **The alignment between the PSSA and the grade-level Assessment Anchors and Eligible Content statements in mathematics is strong, as the data in the following table attest.**

Achieve remapped some items to different Eligible Content statements than that originally specified by the test developers to ensure the content and performance of items matched the statements as closely as possible. Having done so, Achieve found the alignment of Pennsylvania's tests to the Assessment Anchors and Eligible Content statements to be quite strong, as the following summary table demonstrates.

GRADE LEVEL	MAPPED ITEMS Number/Percent	CONTENT CENTRALITY SCORES Number/Percent				PERFORMANCE CENTRALITY SCORES Number/Percent			
		2	1a	1b	0	2	1a	1b	0
3	56	46	1	5	4	44	0	8	4
	100%	82%	2%	9%	7%	79%	0%	14%	7%
5	55	54	0	1	0	46	0	9	0
	100%	98%	0%	2%	0%	84%	0%	16%	0%
8	57	50	0	7	0	48	0	7	2
	100%	88%	0%	12%	0%	84%	0%	12%	4%
11	57	45	0	11	1	50	0	4	3
	100%	79%	0%	19%	2%	88%	0%	7%	5%
TOTAL	100%								

The vast majority of items received scores of “2” for content centrality, indicating strong alignment, while remaining items were typically scored “1b.” Scores of “1b” indicate partial alignment because the related standard addresses more than one concept or topic. Scores of “1a” signal that a standard is written at too high a level of generality to be sure of a given item’s alignment. The fact that only one item at grade 3 received a score of “1a” for content centrality demonstrates that the anchors and related content statements are clear and specific.

Although the match of the performances required by the items to the anchors and related content statements was slightly less (except for grade 11) than the match of content, alignment scores overall were quite good.

- **Pennsylvania’s tests of mathematics in grades 3, 5, 8 and 11 are generally well crafted.**

Achieve uses the “source of challenge” criterion to flag items that are flawed or “unfair” — the source of the challenge of the item lies in something other than the content or performance. Calling attention to flawed items on a test is important because defective items may cause a state to end up with an incorrect perception of precisely which concepts or skills are causing difficulty for students.

There are two ways in which an item merits a source of challenge score of “0” (i.e., has an inappropriate source of challenge): (1) An item is automatically scored “0” for source of challenge if it has been scored “0” for both content and performance centrality. Because it is not aligned to the related standard, it is not a fair item, falling as it does outside the state’s eligible content; (2) an item is scored as “0” if it contains a technical flaw that might lead a student either to get the right answer for the wrong reason or to get the wrong answer but possess the knowledge to answer the item correctly. Examples of common technical flaws Achieve has encountered in mathematics tests from around the nation include having no correct answer or multiple correct answers, misleading graphics, incorrect labels on figures, and items in which the reading is more of a factor than the mathematics. The following table summarizes Achieve’s findings for this criterion.

Source of Challenge Scores

GRADE	NUMBER OF MAPPED ITEMS	SCORE = 1 Number and Percent	SCORE = 0 Number and Percent
3	55	50/91%	5/9%
5	56	52/93%	4/7%
8	57	57/100%	0/0%
11	57	55/96%	2/4%

The data indicate there were few technical flaws noted in the PSSA items, although the state will want to give close attention to grade 3 items, of which reviewers found an appreciable number (9 percent) that had an inappropriate source of challenge. Items were found to be problematic for reasons of trivializing the mathematics, being “generic” and a poor match to the content or performance delineated by the related content statements, or being cast as constructed-response with no “value added” over a multiple-choice format. The grade 5 test also deserves a second look, being that one of its constructed-response items, worth 10 points, was found to have an inappropriate source of challenge.

- **Pennsylvania includes constructed-response items on its mathematics assessments.**

Pennsylvania has made a good decision by including constructed-response items on its tests. Constructed-response items that are well written afford students the opportunity to demonstrate their mathematical understandings in an authentic way and can provide a mechanism for testing aspects of the anchors that are difficult to assess via multiple-choice items. Achieve recommends that the state use its grade 8 constructed-response items, along with their rubrics, as models to inform test developers about the quality the state wants to see in its constructed-response items.

- **Individual grade-level tests have some additional noteworthy positive features.**

Grade 5: The grade 5 test provided excellent coverage of the Pennsylvania anchors and was obviously put together with that in mind. It may be that the process used to check the balance of test items on this particular test was slightly different than that used to conduct a final review of the remaining tests, and if so, this process should be emulated.

Grade 8: As noted above, the constructed-response items at grade 8 are excellent and should serve as exemplars. The items are fair but require thought, and they are informative (neither trivial nor overscaffolded), making them worth students’ time to answer. As a result, they provide critical information to the state regarding student proficiency that cannot be garnered solely through items that are cast in a multiple-choice format.

Grade 11: The grade 11 test’s level of cognitive demand was closer to grade appropriateness than in many states where Achieve has found the tests at high school to be barely distinguishable from those at grade 8.

AREAS FOR IMPROVEMENT

- **Pennsylvania needs to strengthen the geometry content of its tests.**

All of the state's tests were found to have a relatively consistent weakness in geometry. At grades 3, 5 and 8, the level of cognitive demand of items assessing this strand was relatively low. Also, at every grade, the content assessed rarely was central to that grade level and too often was grounded in definitions and little more. The root cause of this weakness may lie in the Assessment Anchors and Eligible Content statements that relate to geometry. When standards suffer from a lack of rigor or precision, or fail to show an increasing trajectory of demand from grade to grade, it is challenging for test developers to compensate for these deficiencies in constructing test questions. In addition, at grade 11 there were too few items assessing geometry, while at grade 8 the set of items assessing geometry was not well balanced, placing too much emphasis on transformation and ordered pair sets.

- **Pennsylvania's tests in mathematics do not consistently match the rigor of the Assessment Anchors and Eligible Content statements.**

Pennsylvania will want to fine tune its tests to ensure they are assessing the more rigorous knowledge and skills described in its anchors and Eligible Content statements. The overall level of cognitive demand of the grade 11 test was better than many high school tests Achieve has reviewed. However, the state should increase the cognitive demand of its tests in part by attending to the issues discussed below.

The state may want to take a close look at items that received scores of "1b" for content or performance centrality. A single Eligible Content statement may address more than one topic — for example, "roots and exponents" — and/or more than one kind of performance — for example, "identify and apply." However, a single item generally assesses only one part of a compound content or performance statement. Such items are flagged in the review process with a "1b" score, signaling that the item is aligned to a part of the statement. Achieve has found that too often test developers target the less demanding content or performance in constructing items, and Achieve's reviewers look across "1b" items to see if that kind of pattern exists.

This was the case for Pennsylvania's assessments. Twelve percent of mapped grade 8 items and 19 percent of mapped grade 11 items were scored as "1b" for content centrality. Reviewers found the test did not do a good job of sampling the more advanced content described in its statements of Eligible Content. Where there were multiple content strands or topics described in Pennsylvania's statements, test questions tended to assess the least-demanding content described in the corresponding statement.

At grades 3, 5 and 8, the state did not evenly sample the more demanding of the performances called for in the Eligible Content statements. Reviewers scored 14 percent of grade 3, 16 percent of grade 5 and 12 percent of grade 8 mapped items as "1b" for performance centrality. Where the Eligible Content statements called for multiple performances, the state too often targeted the least-demanding performance on its tests. In addition, the state may want to look at the progression of the cognitive demand of items (the level of thinking and reasoning required by the student on a particular item) across tests. Reviewers judged the intellectual demand of each item and rated them on a scale of 1–4, summarized as follows:

-
- Level 1 (Recall) Items require the recall of information such as a fact, definition, term or simple procedure.
 - Level 2 (Skill/Concept) Items call for the engagement of some mental processing beyond a habitual response, with students required to make some decisions as to how to approach a problem or activity.
 - Level 3 (Strategic Thinking) Items require students to reason, plan or use evidence.
 - Level 4 (Extended Thinking) Items require complex reasoning, planning, developing and thinking, most likely over an extended period of time. (These items are not generally found on large-scale, on-demand tests.)

The following table shows the balance of intellectual demand of items at each grade level.

MATHEMATICS ASSESSMENTS		LEVEL OF DEMAND				
Grade	# of Items	1	2	3	4	nr*
3	55	20%	68%	4%	0%	9%
5	56	31%	58%	4%	0%	7%
8	57	37%	58%	5%	0%	0%
11	57	26%	65%	5%	0%	4%

*Items scored as “0” for source of challenge are “unfair” and are not reviewed for level of demand.

Grades 5, 8 and 11 contain a higher percent of items at Level 1 than found at grade 3 — not a trend one would expect as students move up the grades. The vast majority of items on all four tests are Level 2 items, and that is appropriate. One would hope to see a higher percentage of Level 3 items at all grades, but especially at grades 8 and 11, which mark important transitions for students in demonstrating readiness for the next level of increasingly abstract mathematics.

- **Each grade-level test in mathematics has particular problematic areas that merit attention.**

Grade 3: As noted previously, the state will want to review items that scored “0” for source of challenge because either the mathematics content or the student’s opportunity to perform was undercut by the item’s design. In addition, a number of items failed to match the state’s expectations for content and/or performance as described in the Assessment Anchors and Eligible Content statements. This grade-level test also contains the weakest of the state’s constructed-response items.

Grade 5: The constructed-response items were disappointing in that the scoring rubrics were confusing. Moreover, one item, worth 10 points, scored “0” for source of challenge. The set of items assessing the Number strand need special attention: (1) too many items assessed the least-demanding performance where the Eligible Content statement called for multiple performances; (2) the items did not provide balanced coverage — negative numbers and prime numbers were overassessed at the expense of other important eligible content.

Grade 8: Too many items assessed the less advanced content or the less demanding performance when the Eligible Content statement contained multiple topics or called for more than one performance.

Grade 11: The grade 11 test is not well balanced. Achieve recommends that the state aim for a distribution of 30 percent for Algebra; 25 percent for Geometry; and 15 percent each for Number, Measurement, and Data Analysis and Probability. To address the content imbalance issues between the Number and Algebra strands, the state may wish to consider pruning the Number anchor statements so there are fewer of them and their total is more in line with the appropriately low emphasis given to Number on the grade 11 test. The state also could redress this imbalance by decreasing Algebra's allotment in the test blueprint and test to about 30 percent. To address content imbalance issues between Algebra and Geometry, Pennsylvania should consider parsing two- and three-dimensional figures, congruence, and similarity in the Geometry Eligible Content statements and then writing items to assess all the Geometry statements.

RECOMMENDATIONS FOR IMPROVEMENT

- ✓ **Raise the rigor of the assessments to ensure they reflect the rigor of the Assessment Anchors and Eligible Content statements.**

The assessments did not evenly sample the more advanced content and performances described by the Eligible Content statements. In addition, the state should review the cognitive demand of its tests across the grades. With the exception of the grade 3 test, the state should prune the number of items with a relatively low level of cognitive demand (Level 1) and increase the number of items requiring more advanced skills (Level 3).

- ✓ **Review the state's purpose(s) for including constructed-response items on the test, and communicate these clearly to the developers.**

Constructed-response items should further the state's ends by providing information regarding student proficiency not readily obtainable from items in a multiple-choice format. The state will want to review closely not only the items but also the rubrics to make sure they are clear and free of internal contradictions.

- ✓ **Consider establishing course requirements and end-of-course tests in Algebra I, Geometry and Algebra II.**

Achieve's recent studies leave little doubt that students need to demonstrate greater proficiency in mathematics and strongly recommend that states phase in a rigorous program of four years of high school mathematics to include Algebra I, Geometry, and Algebra II, as well as data analysis and statistics, as quickly as possible. No state yet has these requirements in place for all students, but three states — Arkansas, Indiana and Texas — are close and require parents who would prefer a less challenging curriculum for their children to opt out deliberately. Achieve is mindful that for students to succeed in such a demanding program, they must enter high school with secure foundational knowledge and skills. For this reason, we currently are at work producing a K–12 set of mathematics standards that would have students college and work ready by the end of grade 12. As Pennsylvania moves forward in implementing a more rigorous program of study, it will want to take advantage of these benchmark standards.

*

*

*

As this report makes clear, Pennsylvania has made great strides in the past two years in strengthening its standards, assessments and instructional program by focusing on essential core content — the durable knowledge and skills in reading and mathematics that students must retain from year to year. The state’s responsiveness to suggestions has led to a set of Assessment Anchors that are aligned with the state’s overarching academic standards; inclusive of important, measurable knowledge and skills at grade levels 3–8 and 11; generally written in clear and specific language; and formatted for ease of use, although in mathematics some changes will not affect assessments until 2007.

Pennsylvania’s decision to focus on Assessment Anchors and Eligible Content statements is not without risk. It makes the state vulnerable to the charge it is reducing rigor, although that clearly is not the state’s intent. To counter the charge, Pennsylvania will need to implement the kind of strong supports mentioned in this report to ensure its K–12 system of instruction and assessments is robust and results in solid gains in what the state’s students know and can do. Consequently, we urge the state to continue work on progression — making sure a rigorous development of content across the grades is readily evident in its Assessment Anchors and Eligible Content statements and buttressed with examples to clarify the level of demand the state expects. Over time, we also urge the state to raise the level of rigor of its standards and tests so its graduates will be prepared for the intellectual demands of postsecondary education and today’s competitive job market. Finally, Achieve encourages Pennsylvania to develop a strong education coalition so the Department of Education, the governor’s office, and the business and higher education communities all speak with one voice, building the kind of support essential for ensuring Pennsylvania’s reform efforts will be effective and its goals realized.

APPENDIX: BIOGRAPHIES

ACHIEVE'S BENCHMARKING STAFF

The following Achieve staff and senior consultants led the analysis and report development for Pennsylvania.

MATTHEW GANDAL **EXECUTIVE VICE PRESIDENT, ACHIEVE**

Matthew Gandal joined Achieve in 1997, shortly after governors and business leaders created the organization. He opened the organization's Washington, DC, office and helped build its programs and services.

As executive vice president, Gandal has senior responsibility for overseeing Achieve's major initiatives. He supervises Achieve's work with states and helps shape the organization's national agenda. He played a lead role in organizing the 1999 and 2001 National Education Summits attended by governors, corporate CEOs and education leaders from across the country.

Gandal has extensive experience reviewing academic standards and education policies in the United States and abroad. He has written dozens of reports and articles on the topic. He also has served on a variety of national and international panels and has helped advise academic standards commissions and legislative bodies in numerous states.

Before joining Achieve, Gandal was assistant director for educational issues at the American Federation of Teachers (AFT), where he oversaw the national organization's work on education standards, testing and accountability. He helped AFT launch a variety of programs and publications designed to support standards-based reform efforts in states and school districts. He was the author and chief architect of *Making Standards Matter*, an annual AFT report evaluating the quality of the academic standards, assessments and accountability policies in the 50 states. He also wrote a series of reports entitled *Defining World Class Standards*, which compared student standards and achievement in the United States with that of other industrialized nations.

Prior to his role with AFT, Gandal served as assistant director of the Educational Excellence Network, an organization founded by Checker Finn and Diane Ravitch. In addition to work on domestic policy issues, he was responsible for directing a series of projects aimed at helping emerging democracies around the world build democratic education systems.

Gandal is a proud graduate of the public school system in the state of Maryland. He earned a bachelor's degree in philosophy from Trinity College in Hartford, CT. He lives in Maryland with his wife and three children.

JEAN SLATTERY
DIRECTOR, BENCHMARKING INITIATIVE, ACHIEVE

Jean Slattery has been with Achieve since 1999 and currently serves as director for the Benchmarking Initiative and lead reviewer in science. She was supervising director of curriculum development and support in Rochester, NY, from 1989 to 1997, with responsibility for overseeing the work of all subject-area directors in the K–12 instructional program. Her earlier responsibilities as a district-level administrator included serving as director of the middle school (1987–89) and junior high (1985–87) programs. During that period, she initiated Teachers as Partners, a peer-coaching staff-development program funded by the Ford and Matsushita (Panasonic) Foundations.

Slattery served as a peer consultant on standards and assessment for the U.S. Department of Education. She also has served as a consultant to the Washington, DC, school district; San Diego Unified School District; a Washington state consortium of rural schools; and the Arkansas and Illinois Departments of Education. She also has worked for the Council for Basic Education on projects involving the Flint Community School District, the Nevada Education Department and the Cleveland Municipal School District.

Slattery received a bachelor's degree in chemistry from Albertus Magnus College, a master's degree in science education from Yale University and a doctorate in science curriculum from the University of Rochester.

JOANNE THIBAUT ERESH
SENIOR ASSOCIATE, ENGLISH LANGUAGE ARTS, ACHIEVE

JoAnne Thibault Eresh is a senior associate at Achieve, where she leads the English language arts aspects of the Standards-to-Standards Benchmarking and Assessment-to-Standards alignment reviews. She taught writing at the university level and English at private and public high schools in St. Louis, MO, and Fitchburg, MA. She began her work in curriculum design and performance assessment in 1979 under Superintendent Richard C. Wallace Jr., and from 1981 to 1994 she was director of the Division of Writing and Speaking for Pittsburgh Public Schools. During that time, she directed The Pittsburgh Discussion Model Project, funded by the Rockefeller Foundation and part of the CHART network, and she later directed the imaginative writing part of the ARTS Propel Project, a joint project with Harvard's Project Zero and the Educational Testing Service. She was the Pittsburgh district coordinator for the New Standards Project and wrote the teachers' guides for the New Standards ELA Portfolios. In 1995, she was one of the original resident fellows at the Institute for Learning at the University of Pittsburgh's Learning Research and Development Center, and she coordinated the New Standards Linking Projects. From 1997 to March 2001, she was the coordinator of staff development in Community District Two in New York City, where she was responsible for the hiring, training and coordination of that district's staff development group.

KAYE FORGIONE
SENIOR ASSOCIATE, MATHEMATICS, ACHIEVE

Kaye Forgione began consulting work with Achieve in 2000 and joined Achieve as senior associate for mathematics in March 2001. Kaye's primary responsibilities are managing Achieve's Standards and Benchmarking Initiatives involving mathematics. Prior to joining Achieve, she served as assistant director of the Systemic Research Collaborative for Mathematics, Science and Technology Education (SYRCE) project at the University of Texas at Austin, funded by the National Science Foundation. Her responsibilities at the University of Texas also included management and design responsibilities for UTeach, a collaborative project of the College of Education and the College of Natural Sciences to train and support the next generation of mathematics and science teachers in Texas.

Before her work at the University of Texas, Forgione was director of academic standards programs at the Council for Basic Education, a non-profit education organization located in Washington, DC. Prior to joining the Council for Basic Education in 1997, she worked in the K–12 arena in a variety of roles, including several leadership positions with the Delaware Department of Education. She began her education career as a high school mathematics teacher and taught mathematics at the secondary and college levels as part of adult continuing education programs.

Forgione received a bachelor's degree in mathematics and education from the University of Delaware, a master's degree in systems management from the University of Southern California, and a doctorate in educational leadership from the University of Delaware.

MARA CLARK
RESEARCH ASSOCIATE, BENCHMARKING INITIATIVE, ACHIEVE

Mara Clark is the research associate for Achieve's Benchmarking Initiative and assists in the coordination of Achieve's state benchmarking work and the production of the initiative's publications. She also contributes to the English language arts Standards-to-Standards benchmarking and Assessment-to-Standards alignment reviews. Before joining Achieve in this capacity, she was with the American Diploma Project (ADP), a joint partnership of Achieve, The Education Trust and the Thomas B. Fordham Foundation. While with the American Diploma Project, she worked closely with postsecondary faculty, high school teachers and business representatives from across the country on the development of ADP's benchmarks for college and workplace readiness. Clark holds a bachelor's degree in English from the University of Dallas in Irving, TX.

CONSULTANTS AND EXPERT REVIEWERS

Achieve relied on the expertise of nationally respected experts in academic content, standards, curriculum and assessment design to inform and conduct the standards benchmarking and alignment of assessments to standards.

ENGLISH LANGUAGE ARTS

ELIZABETH HAYDEL (CO-LEAD REVIEWER WITH JOANNE THIBAUT ERESH)

Elizabeth Haydel was the project manager for Indiana University’s Center for Innovation in Assessment in Bloomington and the project coordinator for the Center for Reading and Language Studies. A graduate of Stanford University with a degree in American studies, she also holds a master’s degree in language education from Indiana University. She is currently an English language arts consultant for the Ohio Department of Education.

Haydel has taught reading and writing to high school students who had failed Indiana’s statewide achievement test and “Reading in the Content Areas” for Indiana University’s language education department. She has co-written a number of reading workbooks for children, including the Steck-Vaughn *Think-Alongs: Comprehending While You Read* program. She has written test passages and items for various state assessment and test preparation programs.

EUNICE GREER

Eunice Ann Greer is a principal research analyst at the American Institutes for Research in Washington, DC. Her work is focused on assessment design and development and the alignment and implementation of standards-based systems of instruction and assessment. She was an associate superintendent for the Illinois State Board of Education, where she directed the Illinois Reads-Statewide Reading Initiative. Prior to that, she was the division administrator for standards and assessment for the Illinois State Board of Education. Greer was instrumental in Illinois’ successful application for a \$37 million dollar Reading Excellence Act Grant from the Department of Education. Under her leadership, Illinois was the first state to receive the Five Star Award for Exemplary Statewide Reading Initiatives from the International Reading Association. Greer also has worked as an assistant professor in the Department of Curriculum and Instruction at the University of Illinois at Urbana-Champaign, as the director of research for an urban middle school reform project at the Harvard Graduate School of Education, and as a literacy assessment coordinator for the University of Illinois’ Center for the Study of Reading.

MATHEMATICS

HAROLD ASTURIAS (LEAD REVIEWER FOR THE STANDARDS REVIEW)

Harold Asturias is the deputy director of mathematics and science professional development at the University of California Office of the President. He provides oversight to the Mathematics Professional Development Institutes (MPDI) and the California Subject Matter Projects (CSMP). Both statewide projects join K–12 teachers with university faculty to improve teacher content knowledge. Previously, he served as the director of the New Standards Portfolio Assessment Project and the Mathematics Unit for New Standards. In that capacity, he led the development team of experts whose efforts, involving many states and more than 1,000 teachers, resulted in the successful production of two assessment systems: the New Standards Portfolio and the Reference Examination. Asturias was a member of the writing group for NCTM's Assessment Standards for School Mathematics. He has extensive experience providing professional development in the areas of standards and assessment in mathematics for teachers in large urban districts (Chicago, Los Angeles, New York City) and small rural districts. Over the past three years, he has focused in the area of designing and implementing professional development for K–12 California mathematics teachers who teach English language learners.

PAM BECK (LEAD REVIEWER FOR THE ASSESSMENT REVIEW)

Pam Beck taught for 10 years in central California public schools before joining the Balanced Assessment team at the University of California at Berkeley (a project funded by the National Science Foundation). This team produced assessment tasks for students from the elementary to high school levels that were published by Dale Seymour under the title *Balanced Assessment for the Mathematics Curriculum*. Since 1994, she has worked at the University of California developing mathematics curriculum and assessment.

During this time, Beck directed the development of a standards-based mathematics examination (the *New Standards Reference Examination*) given at the elementary, middle and high school levels. She helped develop the *New Standards Performance Standards*. She worked as part of the team that wrote *Core Assignments in Mathematics*, published by the National Center on Education and the Economy. During this same period, she provided professional development to numerous and varied districts (including Hanford, CA; Los Angeles Unified; and New York City). She currently directs an NSF-funded project to develop a Web-based task bank. This task bank's purpose is to provide teachers and others with a wide variety of mathematics tasks useful for classroom assessment and indexed for optimum usefulness.

MARY LYNN RATH

Mary Lynn Rath received her bachelor's degree in mathematics from Indiana University at Pittsburgh and her master's in mathematics education from the University of Pittsburgh. She is currently a mathematics specialist in the Division of Instructional Support of the Pittsburgh Public Schools. Her responsibilities include leadership roles in curriculum development, textbook selection, design of alternative assessments, in-service

program design and implementation, and coordination of mathematics programs across levels and schools. She has special responsibility for middle schools. She also is the co-director of the Pittsburgh Reform in Mathematics Education project (PRIME), a K–12 professional development system.

Prior to this position, Raith was a mathematics supervisor (1986–1996) in Pittsburgh and a middle school mathematics specialist in grades 6 through 8 (1970–1986) working with remedial as well as gifted students. She has designed and presented — locally, regionally and at national conferences — sessions on the infusion of algebraic thinking, geometric reasoning, statistics and probability, and problem solving into the K–8 mathematics program. In summer 1987, she was chosen to attend the Michigan State University honors teachers workshop, and since then she has been involved with the implementation, piloting and in-servicing of MSU programs.

She also has been involved with a number of national projects, including the development of both the *New Standards Reference Examination* and the Portfolio project for the middle grades, the Assessment Communities of Teachers project (ACT), and the Alternative Assessment in Mathematics project (A²IM). She also has worked extensively with both NCTM and NCEE on the America's Choice school design and has presented at numerous national conferences.

JOSEPH L. ACCONGIO **PROJECT ADMINISTRATOR, ACHIEVE**

Joseph L. Accongio is a consultant and the former principal and superintendent of the Charter School of Science and Technology in Rochester, NY. He also was the school's director of program development and the primary charter recipient. He has been principal of the Nathaniel Rochester Community School and Thomas Jefferson Middle School, as well as the house administrator of the Discovery Magnet at Frederick Douglass Middle School. In addition, Accongio was a curriculum coordinator and science teacher, chemistry teacher and biology teacher in the Rochester City School District.

Accongio spent a year as director of school services with the Children's Television Workshop, creators of *Sesame Street*, *3-2-1 Contact* and *Square One TV*. He developed a series of teachers' guides for the science and mathematics shows and conducted numerous workshops on using these popular shows in the classroom. He also co-wrote a monograph on science assessment entitled "Classroom Assessment — Key to Reform in Science Education."

He received a doctorate in curriculum planning from the State University of New York (SUNY) at Buffalo, a master's degree in education from SUNY at Brockport, and a bachelor's degree in general sciences from the University of Rochester, NY.

Officers and Board of Directors

Achieve's board of directors is composed of six governors (three Democrats and three Republicans) and six CEOs.

Co-Chairs

Governor Bob Taft
State of Ohio

Arthur F. Ryan, Chairman and CEO
Prudential Financial, Inc.

Vice Chair

Kerry Killinger, Chairman and CEO
Washington Mutual

Board Members

Craig R. Barrett, CEO
Intel Corporation

Governor Ernie Fletcher
Commonwealth of Kentucky

Governor Jennifer Granholm
State of Michigan

Jerry Jurgensen, CEO
Nationwide

Governor Edward G. Rendell
Commonwealth of Pennsylvania

Governor Mike Rounds
State of South Dakota

Edward B. Rust, Jr., Chairman and CEO
State Farm Insurance

Chairman Emeritus

Louis V. Gerstner, Jr., Former Chairman and CEO
IBM Corporation

President

Michael Cohen



Achieve, Inc.
www.achieve.org

1775 Eye Street, NW
Suite 410
Washington, DC 20006
Phone: (202) 419-1540
Fax: (202) 828-0911